



HAL
open science

Analogy in recommendation. Numerical vs. ordinal: A discussion

Nicolas Hug, Henri Prade, Gilles Richard, Mathieu Serrurier

► **To cite this version:**

Nicolas Hug, Henri Prade, Gilles Richard, Mathieu Serrurier. Analogy in recommendation. Numerical vs. ordinal: A discussion. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016) in IEEE World Congress on Computational Intelligence - WCCI 2016, Jul 2016, Vancouver, Canada. pp. 2220-2226. hal-01787372

HAL Id: hal-01787372

<https://hal.science/hal-01787372v1>

Submitted on 7 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is a publisher's version published in : <http://oatao.univ-toulouse.fr/> Eprints
ID : 18962

The contribution was presented at FUZZ-IEEE 2016:

<http://www.wcci2016.org/>

To link to this article URL :

<http://dx.doi.org/10.1109/FUZZ-IEEE.2016.7737969>

To cite this version : Hug, Nicolas and Prade, Henri and Richard, Gilles and Serrurier, Mathieu *Analogy in recommendation. Numerical vs. ordinal: A discussion.* (2017) In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016) in IEEE World Congress on Computational Intelligence – WCCI 2016, 24 July 2016 - 29 July 2016 (Vancouver, Canada).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Analogy in Recommendation. Numerical vs. Ordinal: a Discussion

Nicolas Hug*, Henri Prade*, Gilles Richard*, Mathieu Serrurier*

*Institut de Recherche en Informatique de Toulouse,

CNRS, Université Paul Sabatier, France

Email: {nicolas.hug, henri.prade, gilles.richard, mathieu.serrurier}@irit.fr

Abstract—The paper investigates the use of analogical reasoning for recommendation purposes. More particularly, we address the problem of predicting missing ratings on the basis of known ones. After discussing the differences with another recently experimented approach based on analogical proportions, a new analogical approach is proposed. It relies on the intuition that “the rating of user u for item i is to the rating of user v for item i as the rating of user u for item j is to the rating of user v for item j ”. This leads to algorithms yielding results close to the ones of state-of-the art approaches, when the ratings are regarded as numerical quantities. This is due to the fact that these latter approaches embed an estimation process that is implicitly close to analogy, as discussed in this paper. An analogical approach is also outlined and briefly discussed when the ratings are supposed to have an ordinal meaning only.

I. INTRODUCTION

Recommendation may refer to a variety of problems depending on the information available. One may try to propose items or products on the basis of their descriptions to users whose preferences profiles are known. One may also take advantage of the behavior of other users that are similar to the recommendee. One may also try to predict missing ratings on the basis of known ratings. Exploiting preferences may call for fuzzy set methods; see, e.g. [1], for an early example. Similarity is also a graded notion that underlies case-based reasoning, which can be embedded in a fuzzy rule-based approach and be related to k -nearest neighbor approaches [2], [3], [4].

The recommendation problem considered in this paper is the prediction of missing ratings on the basis of known ratings. We more particularly explore the idea of applying analogical reasoning to this problem. Analogy is used here in terms of analogical proportions, i.e., statements of the form “ a is to b as c is to d ”. In case-based reasoning, situations with known conclusions are put in parallel one by one, with a new pair (situation 0, conclusion 0) where ‘conclusion 0’ is unknown. Then case-based reasoning can be viewed as a particular instance of analogical reasoning since one can say that “conclusion 0 should be to conclusion i as situation 0 is to situation i ”. However there is a more sophisticated way to apply analogy here, namely to state that “(situation 0, conclusion 0) is to (situation 3, conclusion 3) as (situation 2, conclusion 2) is to (situation 1, conclusion 1)”, which requires to put the situation on which one wants to conclude in parallel with three other situations where the corresponding conclusion

is known [5]. Then using a formal model of an analogical proportion [6], [7], and observing that analogical proportions hold on various features describing the four situations, one concludes that “conclusion 1 is to conclusion 2 as conclusion 3 is to conclusion 0” should hold as well, which leads to compute ‘conclusion 0’ from this latter relation.

The idea of applying analogy to recommendation is not entirely new. Thus, Sakaguchi et al. [8] use four-terms analogy in a case-based reasoning style for proposing dishes to users, while three of the authors of the present paper have more recently proposed a 4-(situation, conclusion)-based analogical mechanism for predicting missing ratings on the basis of known ratings [9]. This latter work yielded reasonably good results, but was extremely heavy computationally speaking. In this paper, we investigate a more tractable way of using analogical proportions for solving the same problem. Namely, letting r_{ui} be the rating for item i by user u , we assume that “ r_{ui} is to r_{vi} as r_{uj} is to r_{vj} , where r_{vj} is unknown, while the three other ratings are available. We shall first consider the ratings as numbers, which leads to an estimation process quite close to the one used in Takagi-Sugeno fuzzy rule-based controllers [10] where similarity-based weighted averages are performed. We then more briefly discuss the case where the ratings are only considered as having an ordinal meaning.

The paper is structured as follows. The next section provides the necessary background on the modelling of analogical proportions when features are Boolean and when they are numerical. Section 3 first recalls the previously proposed analogical approach to the prediction of missing ratings which uses the sophisticated mechanism involving four parallel vectors of the (situation, conclusion)-type. Then Section 3 introduces the way analogy is applied in this paper. Section 4 presents the algorithm that exploits this view and reports results of experiments on the Movielens benchmark when the ratings are regarded as numerical quantities. These results are quite close to the ones obtained by state-of-the art approaches. This is due to the fact that the proposed analogical approach appears to be formally very close to the state-of-the art approaches, although the latter do not refer to analogy at all, as revealed by the discussion ending the section. Section 5 outlines an ordinal counterpart to the proposed analogical approach, since it is arguable that ratings have often mainly an ordinal meaning.

II. ANALOGICAL REASONING WITH PROPORTIONS

The following section provides the necessary background on analogical reasoning that will be used throughout this paper.

A. Formal definitions

An analogical proportion “ a is to b as c is to d ” states analogical relations between the pairs (a, b) and (c, d) , as well as between the pairs (a, c) and (b, d) . There are numerous examples of such statements, with which everybody will more or less agree, such as “calf is to cow as foal is to mare”, or “brush is to painter as chalk is to teacher”. However, it is only rather recently that formal definitions have been proposed for analogical proportions, in different settings [11], [12], [13]. For more details, see [7], [14], [15].

It has been agreed since Aristotle time, taking lesson from geometrical proportions, that an analogical proportion T , as a quaternary relation, satisfies the three following characteristic properties:

- 1) $T(a, b, a, b)$ (reflexivity)
- 2) $T(a, b, c, d) \implies T(c, d, a, b)$ (symmetry)
- 3) $T(a, b, c, d) \implies T(a, c, b, d)$ (central permutation)

There are various models of analogical proportions, depending on the target domain. When the underlying domain is fixed, $T(a, b, c, d)$ is simply denoted $a : b :: c : d$. Standard examples are:

- Domain \mathbb{R} : $a : b :: c : d$ iff $a - b = c - d$ iff $a + d = b + c$ (arithmetic proportion)
- Domain \mathbb{R}^n : $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ iff $\vec{a} - \vec{b} = \vec{c} - \vec{d}$. This is just the extension of arithmetic proportion to real vectors. In that case, the 4 vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ build up a parallelogram.
- Boolean domain $\mathbb{B} = \{0, 1\}$:
 $a : b :: c : d$ iff $(a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c)$

In the following, we will be mostly interested in the arithmetic proportions in \mathbb{R} or in \mathbb{R}^n , and will work with analogies between ratings.

B. Using analogical proportion for inference

To understand how one can infer new information on the basis of analogical proportions, we need to define the equation solving process. The equation solving problem amounts to finding a fourth element x to make the incompletely stated proportion $a : b :: c : x$ to hold. As expected, the solution of this problem depends on the target model. For instance, in the case of extended arithmetic proportions, the solution always exists and is unique: $x = b - a + c$. In terms of geometry, this simply tells us that given 3 points, we can always find a fourth one (aligned with, or in the same plan as a, b, c) to build a parallelogram.

The analogical inference principle is, logically speaking, an unsound inference principle, but providing plausible conclusions [16]. It postulates that, given 4 vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ such that the proportion holds on some components, then it should also hold on the remaining ones. This can be stated as (where $\vec{a} = (a_1, a_2, \dots, a_n)$, and $J \subset [1, n]$):

$$\frac{\forall j \in J, a_j : b_j :: c_j : d_j}{\forall i \in [1, n] \setminus J, a_i : b_i :: c_i : d_i} \quad (\text{analogical inference})$$

This principle leads to a prediction rule in the following context:

- 4 vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ are given where \vec{d} is partially known: only the components of \vec{d} with indexes in J are known.
- Using analogical inference, we can predict the missing components of \vec{d} by solving (w.r.t. d_i) the set of equations (in the case they are solvable):

$$\forall i \in [1, n] \setminus J, \quad a_i : b_i :: c_i : d_i.$$

In the case where the items are such that their last component is a label, applying this principle to a new element \vec{d} whose label is unknown leads to predict a candidate label for \vec{d} .

This prediction technique has been successfully applied to classification problems in both Boolean [17] and numerical settings [18], thus suggesting promising results in the recommendation task.

III. ANALOGY AND THE RECOMMENDATION PROBLEM

This section provides some necessary background on the recommendation task, and explores various ideas that can be developed to build an analogical reasoning-based recommender system.

A. Recommendation as prediction of missing ratings

Let us formalize the problem of recommendation. Let U be a set of users and I a set of items. For some pairs $(u, i) \in U \times I$, a rating r_{ui} is supposed to have been given by u to express if he/she likes or not the item i . R denotes the set of all known ratings. Let U_i be the set of users that have rated item i , and I_u is the set of items that user u has rated. I_{uv} defines the set of items rated by both users u and v . The ultimate goal of a recommender system is to provide relevant and personalized recommendations of items to users, and this is usually done by trying to predict users’ ratings for any item in the the system. Note that users and items can play symmetrical roles: indeed, one can see the recommendation problem as recommending items to users or as recommending users to items. In the following, we chose the first view which we find more intuitive for the reader.

The two main families of recommender systems are *content-based* methods where some meta data describing users and items are used, and *collaborative filtering* methods, much more popular, where predictions are computed by taking into account the *social environment* of users which is usually modeled by the ratings they gave. Collaborative techniques are the one we are interested in here, as they have shown to outperform the content-based ones.

It is quite common that ratings belong to $[0, 1]$ or to $[1, 5]$, while 1 is the worst rating and 5 meaning a strong preference. It is not always clear whether this rating scale should be interpreted as purely numerical, or more like an ordinal scale when it comes to develop prediction algorithms. This is a question that we will address throughout this paper.

When ratings are treated as numerical quantities, the two most used performance evaluation metrics are MAE (Mean

Absolute Error) and RMSE (Root Mean Squared Error), and are usually computed using cross validation:

$$\text{MAE} = \frac{1}{|R|} \cdot \sum_{r_{ui}} |\hat{r}_{ui} - r_{ui}| \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{|R|} \cdot \sum_{r_{ui}} (\hat{r}_{ui} - r_{ui})^2} \quad (2)$$

They both evaluate how close predictions are from their true values, RMSE being much more penalizing over big errors.

B. Analogical proportions between users

Using analogical reasoning for recommendation as been studied in [9]. Authors strictly follow the analogical inference principle described in section II-B for making predictions, using analogical proportions between users.

The main idea is that if an analogical proportion stands between four users a, b, c, d , meaning that for each item j that they have commonly rated, the analogical proportion $r_{aj} : r_{bj} :: r_{cj} : r_{dj}$ holds, then it should also hold for an item i that a, b, c have rated but d has not (i.e. r_{di} is the missing component). This leads us to estimate r_{di} as the solution $x = \hat{r}_{di}$ of the following analogical equation:

$$r_{ai} : r_{bi} :: r_{ci} : x.$$

Given a pair (u, i) such that $r_{ui} \notin R$ (i.e. there is no available rating from user u for item i), the main procedure is as follows:

- 1) find the set of 3-tuples of users a, b, c such that an analogical proportion stands between a, b, c , and u and such that the equation $r_{ai} : r_{bi} :: r_{ci} : x$ is solvable.
- 2) solve the equation $r_{ai} : r_{bi} :: r_{ci} : x$ and consider the solution x as a candidate rating for r_{ui} .
- 3) set \hat{r}_{ui} as an aggregate of all candidate ratings.

This technique has shown to be not too far from basic collaborative filtering approaches [9], but suffers of its inherent cubic complexity which makes it impossible to look for every possible 3-tuples of users, thus compromising the prediction accuracy.

C. Pairwise analogy between clones

Considering analogy between four users has shown to be computationally intensive, thus not really suitable for recommendation purposes, where time is a highly critical dimension. Yet, other forms of analogy can be addressed in the recommendation task, based on the observation that some users may be more inclined to give good (or bad) ratings than others. Indeed, ratings are in no way absolute and greatly depend on the subjective appreciation each user has about the rating scale. In the $[1, 5]$ scale for example, two users u and v might semantically agree on an item i describing it as *bad*, but there is a chance that this agreement is not perfectly reflected in the ratings: u might have rated i with $r_{ui} = 1$ and v with $r_{vi} = 3$, simply because from v ' point of view 3 is a *bad* rating, while for u a rating of 3 would simply mean *decent* or *good enough*. In the following, we refer such users that *semantically* agree on their common items (but not necessarily

numerically) as *clones*, as illustrated in Figure 1. Please note that the word *clone* is not used here to mean *strictly identical*, but more in the sense that two clones are two users following parallel paths.

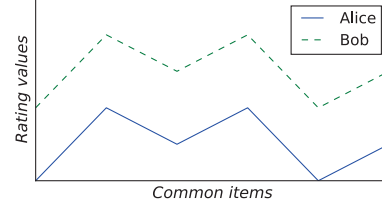


Fig. 1. Bob is a perfect clone of Alice.

It is obvious that in collaborative filtering, clones are of great interest when it comes to predict a user's ratings, and yet the information they provide is often discarded. The principle underlying the analogical clone-based view is the following: for predicting a missing rating for u we not only look at its nearest neighbors, but also to those v whose rating are such that $r_{ui} = r_{vi} + t_{vu}$ where t_{vu} is a more or less constant *correction term* that can be either positive or negative.

In the next two sections, we investigate this idea of a clone-based prediction, first when ratings are viewed as numerical quantities in section IV, and then when they have an ordinal meaning only in section V.

IV. RATINGS AS NUMERICAL QUANTITIES

In the following, we define $C_i(u)$ as the set of users that are clones of u and that have rated item i . From the previous definitions, one can easily derive a very general collaborative filtering framework for predicting a user's rating by taking into account its clones:

$$\hat{r}_{ui} = \text{aggregation}(r_{vi} + t_{vu}), \quad \forall v \in C_i(u),$$

where t_{vu} is a *correction term* that we need to add to v 's ratings so that they correspond to those of u . We clearly have a generalization of the k -NN approach, which we could write as:

$$\hat{r}_{ui} = \text{aggregation}(r_{vi} + t_{vu}), \quad \forall v \in \{v \in C_i(u) | t_{vu} = 0\}.$$

Following this general framework, one can construct a great variety of algorithms with various level of complexity. In the next subsections, we propose a very straightforward algorithm, and a more efficient one.

A. A straightforward prediction algorithm

In its most simple form, a user v can be considered to be a t -clone of u if the ratings of v differ from those of u from a constant t :

$$v \in t\text{-}C(u) \iff \forall i \in I_{uv}, r_{vi} = r_{ui} + t. \quad (3)$$

From then on, computing \hat{r}_{ui} amounts to finding all the users v that satisfy this criteria, and computing an aggregation of their rating for i , which can simply be a mean. We implemented

this basic algorithm described by algorithm 1, and referred to as *Bruteforce*.

Algorithm 1 *Bruteforce*

Input: A set of known ratings R , a user u , an item i such that $r_{ui} \notin R$.

Output: \hat{r}_{ui} , an estimation of r_{ui}

Init:

$C = \emptyset$ // list of candidate ratings

for all users $v \in U_i$ **do**

for all t **do**

if $v \in t\text{-Clones}(u)$ **then**

$C \leftarrow C \cup \{r_{vi} + t\}$ // add x as a candidate rating

end if

end for

end for

$\hat{r}_{ui} = \text{aggr } x$
 $x \in C$

Of course, one may want to relax the definition of a t -clone, as the current one is too strict and only very few users will satisfy this criteria. In our implementation, we chose the following condition:

$$v \in t\text{-}C(u) \iff \sum_{i \in I_{uv}} |(r_{ui} - r_{vi}) - t| \leq |I_{uv}|.$$

This amounts to accept v as a t -clone of u if on average, $r_{ui} - r_{vi}$ is equal to t with a margin of 1.

The values of t clearly depend on the rating scale. The dataset on which we tested our algorithms use the $[1, 5]$ interval, so possible values for t that were considered are integer values between $[-4, 4]$.

This is obviously a very rough algorithm, to which one could point out numerous flaws, but its purpose is to show that even such a basic clone-based approach can lead to better results than a basic neighborhood method.

B. Modeling clones with the similarity measure

Another option to consider clones is to use the well known neighborhood-based formula, and capture their effect inside an appropriate similarity measure. The general neighborhood formula is as follows [19]:

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} r_{vi} \cdot \text{sim}(u, v)}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)},$$

where $N_i^k(u)$ is the set of the k nearest neighbors of u that have rated i . So, we move from a crisp view of the set of clones to a fuzzy one. In fact, the above formula looks very similar to the interpolation principle underlying Takagi-Sugeno fuzzy controller where similarity degree is viewed as a fuzzy membership grade [10].

The above formula is commonly used with classical similarity metrics such as Pearson or cosine similarity, or inverse of MSD (Mean Squared Difference, which is a distance). However, these similarities are not plainly satisfactory when it comes to clones. Indeed with these metrics, two users are

considered to be close if their common ratings are often the same, but two perfect clones u and v with a significant correction term t_{vu} would be considered as far from each other, thus involving a loss of information.

A simple choice to measure how two users relate as clones can be the following:

$$\text{Clone_dist}(u, v) = \frac{1}{|I_{uv}|} \cdot \sum_{i \in I_{uv}} ((r_{ui} - r_{vi}) - \mu_{uv})^2$$

where μ_{uv} is the mean difference between ratings of u and v :

$$\mu_{uv} = \frac{1}{|I_{uv}|} \sum_{i \in I_{uv}} (r_{ui} - r_{vi}).$$

One can understand this distance in two ways:

- it can be regarded as the variance of the difference of ratings between u and v ,
- or it can be regarded as a simple MSD measure ($\text{MSD}(u, v) = \frac{1}{|I_{uv}|} \cdot \sum_{i \in I_{uv}} (r_{ui} - r_{vi})^2$) to which the mean difference of ratings between u and v has been subtracted.

As our measure *Clone_dist* is a distance, it is necessary to transform it into a similarity measure. Common choice is to take its inverse (while accounting for zero division): $\text{Clone_sim}(u, v) = \frac{1}{\text{Clone_dist}(u, v) + 1}$.

Once we know how to find the clones of a user, it is a simple matter to output a prediction using the classical neighborhood approach:

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} (r_{vi} + \mu_{uv}) \cdot \text{sim_clone}(u, v)}{\sum_{v \in N_i^k(u)} \text{sim_clone}(u, v)}.$$

This algorithm will be referred to as *CloneA*. For the sake of completeness, we also tried the same formula but with a more basic similarity metric that does not care about clones: MSD. This algorithm is referred to as *CloneB*.

C. Current practices in neighborhood-based methods

A simple and efficient formula using neighborhood technique, popularized by [20] is the following:

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{v \in N_i^k(u)} (r_{vi} - b_{vi}) \cdot \text{sim}(u, v)}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)}.$$

It is based on a simple k -NN approach, where are added the b_{ui} terms, called *baselines*: $b_{ui} = \mu + b_u + b_i$. μ is the global mean of all ratings in R . The b_u term is intended to capture users propensity to give ratings higher or lower than the global mean μ , and the same goes for items with b_i : some items tend to be rated higher than others. Baselines are computed by solving a least squares problem:

$$\min_{b_u, b_i} \sum_{r_{ui} \in R} (r_{ui} - (\mu + b_u + b_i))^2,$$

which can be achieved efficiently by stochastic gradient descent, or alternating least squares.

Among recommended similarity metrics, this one is of particular interest:

$$\text{sim}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - b_{ui}) \cdot (r_{vi} - b_{vi})}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - b_{ui})^2} \cdot \sqrt{\sum_{i \in I_{uv}} (r_{vi} - b_{vi})^2}}.$$

It is simply a Pearson correlation coefficient, except that instead of centering ratings by their means, they are centered with the baseline predictors. An intuitive and illuminating way to look at this algorithm as a whole is to see that it conceptually follows these steps:

- 1) Compute R' , the set of all ratings normalized by the corresponding baseline: $r'_{ui} = r_{ui} - b_{ui}$. R' can be regarded as the set where all ratings are given from the same frame of reference, thus discarding any bias. In R' , ratings can then be considered as absolute.
- 2) Using R' , compute similarities between users using the cosine similarity (the cosine similarity is the same as the Pearson correlation coefficient, except that quantities are not centered).
- 3) Output a prediction using the basic k -NN formula. As this prediction belongs to the same space of R' where ratings have no bias, it needs to be transposed back to the space of R (for performance evaluation purposes).

In what follows, this algorithm is referred to as k -NNbsl.

It is very clear that the use of the baseline predictors is motivated by the same reasons one would want to consider clones in a rating prediction algorithms. This means that k -NNbsl implicitly takes the idea of clones into account, and thus a form of analogical reasoning. Differences and resemblances of these two approaches are discussed in the next section.

D. Experiments and discussion

We evaluated the performance of the aforementioned algorithms in terms of MAE and RMSE on two datasets, the movielens-100K and movielens-1M datasets¹, containing 100,000 and 1M ratings respectively. Results are shown in tables I and II and where calculated using 5-folds cross-validation. For each of these algorithms, the number of neighbors or clones used to output a prediction is $k = 40$, except for the bruteforce algorithm where the number of clones can not be controlled.

TABLE I
PERFORMANCE OF ALGORITHMS ON THE MOVIELENS-100K DATASET

	k-NN	Bruteforce	Clone A	Clone B	k -NNbsl
RMSE	.9763	.9461	.9353	.9311	.9338
MAE	.7705	.8576	.7327	.7321	.7337

It is very clear that even a very straightforward approach of the clone-based recommendation principle significantly outperforms the most basic k -NN algorithm. It is however a lot heavier to compute, thus not very suitable for real world

¹<http://grouplens.org/datasets/movielens>

TABLE II
PERFORMANCE OF ALGORITHMS ON THE MOVIELENS-1M DATASET

	k-NN	Bruteforce	Clone A	Clone B	k -NNbsl
RMSE	.9216	.	.8996	.8969	.8879
MAE	.7252	.	.7057	.7050	.7005

recommendation purposes (its performances on the Movielens-1M dataset simply could not be computed). The two other clone-based algorithms however, have the exact same complexity of any k -NN-based algorithm which is a significant improvement from the algorithm described in section III-B.

Surprisingly enough, out of the two Clone algorithms, it is the one that does not care about clones in its similarity measure that achieves the best results. This might be due to the fact that in the neighborhood based on MSD, μ_{uv} is necessarily small and thus easier to estimate in a statistical significant way.

Performances of the Clone algorithms are close to those of the state of the art k -NNbsl algorithm. It is however important to understand that these algorithms differ on the following points:

- The Clone algorithms do not address item bias, which is a significant drawback. It may not be unreasonable to believe that incorporating item bias in the prediction would lead to better results.
- There is a subtle yet meaningful difference of interpretation between the biases induced by both algorithms. In the clone algorithm, biases are all pairwise, meaning that they involve two users, and they are computed on items that both users have rated. As for the k -NNbsl algorithm, there is no such thing as a pairwise bias. Bias for a given user is computed using only its own ratings, and is a result of a global optimization problem involving the global mean of all ratings, which means that every single rating in R has an impact on the bias.
- On the biggest dataset (Movielens-1M), the k -NNbsl algorithm appears to achieve better accuracy than the other algorithms, while this is not the case for the small dataset. A possible explanation is that as baselines are computed on the whole training set, they tend to capture most of the noise when the training set gets bigger, thus improving accuracy compared to more heuristic-based approach.

It should also be noted that in fact, it is recommended to perform a shrinkage on the similarity measure of algorithm k -NNbsl, in order to take into account the number of common items between two users: the more items they share, the more confident we are when computing their similarity [20]. Such an approach can improve significantly both RMSE and MAE of the algorithm. Similarly, in the clone-based approach, it might be of interest to discount clones that rely on a too small number of common items.

V. TOWARDS AN ORDINAL VIEW OF RATINGS

We may wonder if one can devise a counterpart of the numerical clone-based approach, which would be compatible

with an ordinal view of the ratings. Indeed, an extreme way for unbiasing and comparing two sets of ratings is to forget about their numerical values, and only consider their rankings. The idea of viewing ratings in an ordinal manner has been advocated in [21]. In this section, we discuss an ordinal counterpart of the analogical approach previously presented. Analogical reasoning with ordinal data has first been proposed in [22], yet with a different concern.

A. An algorithm for rank prediction

Indeed the idea that “the rating of user u for item i is to the rating of user v for item i as the rating of user u for item j is to the rating of user v for item j may be understood as well in an ordinal manner. This leads to state that “the relative ranking of item i among the ratings given by user u is to the relative ranking of item i among the ratings given by user v as the relative ranking of item j among the ratings given by user u is to the relative ranking of item j among the ratings given by user v .”

This means that we need to compare the rankings given by two users u and v on their common items. In the following, ρ_{ui} denotes the relative ranking of item i out of all the items rated by u . Our goal is to estimate all values of ρ_{ui} , for any user and any item. The main steps of a possible algorithm is as follows:

- 1) Compute similarities between users, based on their rankings. A very popular similarity ranking measure is the Spearman’s rank correlation coefficient, or Spearman’s rho.
- 2) Compute an estimated rank $\hat{\rho}_{ui}$ as an aggregation of all the rankings ρ_{vi} extracted from the k nearest neighbors (using Spearman’s rho as similarity):

$$\hat{\rho}_{ui} = \frac{\sum_{v \in N_i^k(u)} \rho_{vi} \cdot \text{sim}(u, v)}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)}.$$

This is obviously very similar to the neighborhood approach described in section IV-B, but instead of predicting a rating, we output a predicted rank. This approach is denoted as *RankAnlg*.

B. Experiments

We evaluated the performance of our algorithm and compared it to other previously described approaches, using the exact same evaluation protocol as in section IV-D. The Movielens-1m dataset was not benchmarked, as our algorithm is too computationally intensive.

RMSE and MAE are good measure for evaluation rating prediction accuracy, but are not suitable when it comes to evaluate rankings. A better measure is the Fraction of Concordant Pair, which evaluates the probability that given any two items i and j rated by any user u , the system has correctly estimated whether u prefers i over j or the inverse. To compute the FCP, we need to intermediate measures. c_u defines the number of concordant pairs for user u , and d_u its number of discordant pairs. The FCP is then computed over all users as the proportion of concordant pairs.

$$c_u = \{(i, j) \in I^2 \text{ s.t. } \hat{r}_{ui} > \hat{r}_{uj} \text{ and } r_{ui} > r_{uj}\}$$

$$d_u = \{(i, j) \in I^2 \text{ s.t. } \hat{r}_{ui} \geq \hat{r}_{uj} \text{ and } r_{ui} < r_{uj}\}$$

$$FCP = \frac{\sum_{u \in U} c_u}{\sum_{u \in U} c_u + \sum_{u \in U} d_u}$$

Note that \hat{r}_{ui} here may represent either a rating prediction or a ranking prediction $\hat{\rho}_{ui}$.

Results are reported in table III.

TABLE III
PERFORMANCE OF ALGORITHMS ON THE MOVIELENS-100K DATASET
(RANKING EVALUATION)

	RankAnlg	k-NN	k-NNbsl
FCP	.7063	.7096	.7163

Unfortunately, even a basic algorithm that was not designed for ranking prediction performs better in terms of FCP. To explain this difference, one may look at the distribution of average support over all the predictions, as shown on figure 2. Between two users u and v , the support is defined as the number of common items ($|I_{uv}|$), which was used to compute the similarity between u and v . For a given prediction \hat{r}_{ui} , the average support is the average of all the supports $|I_{uv}|$ over all users $v \in N_i^k(u)$.

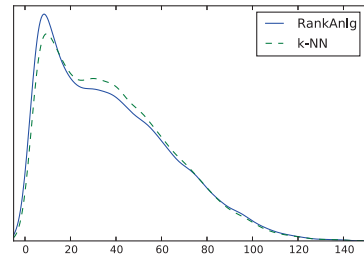


Fig. 2. Distribution of average support.

The use Spearman’s rho tends to provide with neighbors that have smaller support, thus leading to a less significant and less accurate estimation of the neighborhood, which may explain the differences in performance.

VI. CONCLUSION

This paper has provided a discussion on different ways for applying analogical reasoning to the prediction of ratings. After reporting a recent attempt where analogical proportions were built from 4-tuples of users, a computationally simpler approach is presented in this paper based on the idea that “the rating of user u for item i is to the rating of user v for item i as the rating of user u for item j is to the rating of user v for item j .” This agrees with the transitive nature of the underlying analogical modeling. We have shown that this may apply to

a quantitative view of ratings as well as to an ordinal view. Results obtained in the case of the quantitative view remain close to the ones of state-of-the-art approaches, which can be retrospectively reinterpreted in an analogical way.

The idea behind the use of analogies is to go beyond the classical neighborhood to extract relevant information. However, this approach has a cost as it is more difficult to statistically validate the analogical link between users (or items). This is especially true in the ordinal case. Indeed, ordinal analogies tend to select users with a small common support, because it is easy to have the same ranking despite the fact this is not statistically relevant.

In the specific case of MovieLens dataset, a large majority of users seem to have a lot of close neighbors (in the classical sense) from which useful information can be extracted. In that case, examples for which analogical links bring more information than simple neighbors are quite rare. It should not come as a surprise that the pure analogical approach does not bring better results than standard approaches in this dataset.

The analogical approach might be advantageous in the case of low density dataset (i.e. when the set of close neighbors is small). In the same way, we might also think of combining the analogical approach with the classical one provided we are able to detect, for every prediction, which method is statistically the most relevant.

Formalizing analogical reasoning provides tools for extrapolation. This can be done in different ways as shown in this paper, depending on what basis we try to extrapolate. Another issue is to wonder about what we try to extrapolate. Thus, regarding recommendation, one might think of also using analogical reasoning to create configurations describing new items that may plausibly please users.

REFERENCES

- [1] P. Perny and J.-D. Zucker, "Preference-based search and machine learning for collaborative filtering: the "film-conseil" movie recommender system," *Information, Interaction, Intelligence*, vol. 1, no. 1, pp. 9–48, 2001.
- [2] D. Dubois, E. Hüllermeier, and H. Prade, "Fuzzy set-based methods in instance-based reasoning," *IEEE Transactions on Fuzzy Systems*, vol. 10, pp. 322–332, 2002.
- [3] E. Hüllermeier, D. Dubois, and H. Prade, "Model adaptation in possibilistic instance-based reasoning," *IEEE T. Fuzzy Systems*, vol. 10, no. 3, pp. 333–339, 2002.
- [4] D. Dubois, E. Hüllermeier, and H. Prade, "Fuzzy methods for case-based recommendation and decision support," *J. Intell. Inf. Syst.*, vol. 27, no. 2, pp. 95–115, 2006.
- [5] D. Dubois, H. Prade, and G. Richard, "Case-based and analogical reasoning," in *Festschrift for Ramon Lopez de Mantaras 60 years birthday* (C. Sierra, ed.), pp. 59–77, Barcelona: Artificial Intelligence Research Institute, 2012.
- [6] L. Miclet and H. Prade, "Logical definition of analogical proportion and its fuzzy extensions," in *Proc. Annual Meeting of the North Amer. Fuzzy Information Proc. Soc. (NAFIPS), New-York, May, 19-22, IEEE, 2008*.
- [7] H. Prade and G. Richard, "From analogical proportion to logical proportions," *Logica Universalis*, vol. 7(4), pp. 441–505, 2013.
- [8] T. Sakaguchi, Y. Akaho, K. Okada, T. Date, T. Takagi, N. Kamimaeda, M. Miyahara, and T. Tsunoda, "Recommendation system with multi-dimensional and parallel-case four-term analogy," *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC'11)*, pp. 3137 – 3143, 2011.
- [9] N. Hug, H. Prade, and G. Richard, "Experimenting analogical reasoning in recommendation," in *Proc. 22nd Int. Symp. on Foundations of Intelligent Systems (ISMIS'15), Lyon, Oct. 21-23* (F. Esposito, O. Pivert, M. Hacid, Z. W. Ras, and S. Ferilli, eds.), vol. 9384 of LNCS, pp. 69–78, Springer, 2015.
- [10] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modelling and control," *IEEE Trans. on Systems, Man, Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.
- [11] F. Yvon and N. Stroppa, "Formal models of analogical proportions," Tech. Rep. D008, Ecole Nationale Supérieure des Télécommunications, Paris, 2006.
- [12] Y. Lepage, "De l'analogie rendant compte de la commutation en linguistique." Habilit. à Diriger des Recher., Univ. J. Fourier, Grenoble, 2003.
- [13] L. Miclet and H. Prade, "Handling analogical proportions in classical logic and fuzzy logics settings," in *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), Verona*, pp. 638–650, Springer, LNCS 5590, 2009.
- [14] H. Prade and G. Richard, "Analogical proportions and multiple-valued logics," in *Proc. 12th Eur. Conf. on Symb. and Quantit. Appr. to Reasoning with Uncertainty (ECSQARU'13), Utrecht, Jul. 7-10* (L. C. van der Gaag, ed.), LNAI 7958, pp. 497–509, Springer, 2012.
- [15] H. Prade and G. Richard, "Homogenous and heterogeneous logical proportions," *IfCoLog J. of Logics and their Applications*, vol. 1, no. 1, pp. 1–51, 2014.
- [16] H. Prade and G. Richard, "Reasoning with logical proportions," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010* (F. Lin, U. Sattler, and M. Truszczynski, eds.), AAAI Press, 2010.
- [17] S. Bayouduh, L. Miclet, and A. Delhay, "Learning by analogy: A classification rule for binary and nominal data," *Proc. Inter. Joint Conf. on Artificial Intelligence IJCAI07*, pp. 678–683, 2007.
- [18] H. Prade, G. Richard, and B. Yao, "Enforcing regularity by means of analogy-related proportions—a new approach to classification," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp. 648–658, 2012.
- [19] F. Ricci, L. Rokach, B. Shapira, and P. Kantor, eds., *Recommender Systems Handbook*. Springer, 2011.
- [20] Y. Koren, "Factor in the neighbors: Scalable and accurate collaborative filtering," *ACM Trans. Knowl. Discov. Data*, vol. 4, pp. 1:1–1:24, Jan. 2010.
- [21] Y. Koren and J. Sill, "Ordrec: An ordinal model for predicting personalized item rating distributions," in *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, (New York, NY, USA)*, pp. 117–124, ACM, 2011.
- [22] N. Barbot and L. Miclet, "La proportion analogique dans les groupes. Applications à l'apprentissage et à la génération," in *Proc. Conf. francophone sur l'apprentissage artificiel (CAP), Hammamet, Tunisia, 2009*.