Collective Recommendations

Haukur Jónsson, Silvan Hungerbühler, Grzegorz Lisowski, Max Rapp

Abstract

We tackle the practical problem of finding a good mechanism to recommend a collective set of *news items* to a group of media consumers with possibly very disparate individual interest in the available items. For our analysis, we adapt a formal framework from voting theory in *Computational Social Choice* to the media setting in order to compare the performance of five recommendation mechanisms with respect to desirable properties of recommended sets.

1 Introduction

How to balance the media's core function of providing news that is relevant to society at large against the increasing economic necessity of offering an individually tailored product? News media face a dilemma: Either submit to highly personalized news feeds on online social media networks that drive political fragmentation, partisanship and contribute to the erosion of society's commonly accepted factual base; or risk losing disgruntled readers, who feel that "their" issues are inadequately represented in the mainstream media, to less reliable Internet news outlets.

The present paper takes a step towards addressing this problem by designing and testing a number of recommendation mechanisms for news articles. Based on the tastes consumers have for particular news items, the mechanism constructs a collection of essential articles for the entire group. All that is needed for the mechanism to work is an ordering of the news items from first to last for each individual. The way in which this preference ordering is elicited from the individual is left open; depending on the concrete application, the data can be thought of as explicitly provided by the consumers or gathered by data mining techniques.

Naturally, there are certain properties one would expect such a collection of essential articles to have. The total length of recommended articles for a newspaper's title page, for example, should not exceed its character limit which relates to a problem of making collective choices with a restricted budget (Lu & Boutilier, 2011). Likewise, there are relations between the essential articles and the rankings by the individuals one would like to see respected by a recommendation mechanism. For instance, if all consumers detest a certain news item, then it should certainly not be featured in the essential collection instead of another item much liked by everybody.

This paper aims at better understanding of collective recommendation mechanisms in media settings by formally studying the interaction of mechanisms and properties of their recommendation. We employ formal tools provided by *Social Choice Theory* to analyze benefits and drawbacks of various possible ways to determine a set of essential news items for a group, given each member's individual preferences over said items. We proceed by proving a small number of theorems and running simulations to estimate empirical relationships where proofs are unobtainable.

The paper is structured as follows: In Section 2 we provide the formal definition of the recommendation problem as we want to study it. In Section 3 we formally present desirable properties a collection of recommended articles ought to have. In Section 4 we propose five rules for the task of turning individual preferences into a single recommendation. Section 5 presents theorems. Section 6 contains the methodology and presentation of our simulation results, while Section 7 concludes.

2 Formal Framework

This section specifies the formal framework we use. There is a set of *news items* $A = \{a_1, \ldots, a_m\}$, a subset of which are the *recommended items* $W \subseteq A$ for a group of *consumers* $N = \{n_1, \ldots, n_n\}$. Each consumer has preferences over A represented by a strict, total order $\succ_i$ for $i \in \{n_1, \ldots, n_n\}$. Let $\mathcal{L}(A)$ be the set of such orders. Then the preference orders of a set of consumers N over news items A form a *profile of preferences* $\mathcal{R} \in \mathcal{L}^n$. We put $\mathcal{R} + \mathcal{R}'$ to denote the concatenation of two profiles with disjoint voter sets. Likewise, for any $n \in \mathbf{N}$, $n\mathcal{R}$ denotes the concatenation of n identical copies of $\mathcal{R}$.

Each item in A is assigned a specific *cost* by a function $C : A \to \mathbb{N}$ and *utility for a consumer i* by a function $u_i : \mathcal{L}^n \times N \times A \to \mathbb{N}$. We denote by $C(A)$ the image of A under C. We extend these functions to be also defined on sets: The cost of a recommendation set $W \subseteq A$ is given by $C(W) = \sum_{w \in W} C(w)$; its *total utility* by $u(W) = \sum_{a \in W} u(a)$ where $u(a) = \sum_{i=1}^{n} u_i(a)$. We follow Lu and Boutilier (2011, 4) in deriving pseudo-utilities from readers' preference orders. For present purposes we used the Borda score, that is u_i outputs

the value $m-1$ for consumer i's top item, $m-2$ for the second one and so forth - however, nothing hinges on this choice. Depending on the context, the cost could be interpreted as the time it takes to read an article, the cognitive resources it takes a consumer to digest it or simply character length. Finally, as these resources are limited we assume a *budget* $B \in \mathbb{N}_{\geq 0}$.

The recommendation rule then is a function from profiles, costs and the budget to recommended items: $F : \mathcal{L}^n \times \mathbb{N}^{|A|} \times \mathbb{N}_{\geq 0} \to 2^A$.

3 Axioms and Desirable Properties & Dimensions of Performance

We now describe a series of binary properties - *axioms* - and continuous *metrics of performance* a recommendation set should fulfill.

3.1 Axioms

We adapted most of the axioms suggested by Elkind, Faliszewski, Skowron, and Slinko (2017) for the k-multiwinner case to our budgeted setting:

- **Non-Imposition:** For any set of alternatives A and any $W \subseteq A$ such that $C(W) = B$ there is a profile $\mathcal{R}$ such that $F(\mathcal{R}, C(A), B) = W$.

- **Consistency:** Given two profiles $\mathcal{R}_1$ and $\mathcal{R}_2$, a budget B and a voting rule F, F satisfies *consistency* if whenever $F(\mathcal{R}_1, B) \cap F(\mathcal{R}_2, B) \neq \emptyset$ then $F(\mathcal{R}_1, B) \cap F(\mathcal{R}_2, B) \subseteq F(\mathcal{R}_1 + \mathcal{R}_2, B)$.

- **Homogeneity:** For any profile $\mathcal{R}$ and any $n \in \mathbb{N}$: $F(n\mathcal{R}, B) = F(\mathcal{R}, B)$

- **Monotonicity:** Given any profile $\mathcal{R}$ and any $a \in A$ such that $a \in W$, where $W \in F(\mathcal{R}, B)$, for any profile $\mathcal{R}'$ obtained from $\mathcal{R}$ by raising a by one position in one voter's preference order we have: (1) $a \in W'$, where $W' \in F(\mathcal{R}', B)$ and (2) if a was directly below an option $b \notin W$. then $W \in F(\mathcal{R}', B)$

- **Committee-Monotonicity:** For any profile $\mathcal{R}$, $W \in F(\mathcal{R}, B)$ and $\epsilon \in \mathbb{R}$ there is a set W' such that $W' \in F(\mathcal{R}, B + \epsilon)$ and $W \subseteq W'$.

- **Unanimity:** For any profile $\mathcal{R}$: if all consumers $n \in N$ rank the same set of options W such that $C(W) = B$ on top, then $W \in F(\mathcal{R}, B)$.

We decided not to consider the axioms *Consensus Committee* and *Solid Coalitions* since they have no clear equivalent in the budgeted setting. Likewise, we decided to drop the *Fixed Majority* axiom because almost all rules Elkind et al. (2017) consider in the k-multiwinner setting fail to satisfy Fixed Majority. We, therefore, suspected that we would obtain a similar result and deemed it more interesting to assess a criterion of Condorcet consistency that comes in degrees. The purpose of these axioms is to some extent covered by the General Threshold (GT) and Majority Support (MS) axioms described in the next section. In addition, they enable easy implementation of a recommendation rules corresponding to them in a way described in Section 4.

3.2 Metrics of Performance

In addition to the binary axioms, we designed three continuous metrics of performance specifically adjusted to our setting.

Utility Maximization (AUM)
Let $\mathcal{W}_\mathcal{B}$ be the set of all elements of $\mathcal{P}(A)$ s.t. $C(W) \leq B$. A recommendation set W satisfies the axiom of budgeted Utility Maximization (AUM) iff:

$$W = \underset{W' \in \mathcal{W}_B}{\arg\max}(u(W'))$$

The motivation behind this property is that, arguably, a recommendation set should get consumers the highest possible payoff. Even if it does not, it is of interest to assess how far away it is from the optimum.

General Threshold (GT)
Let $N_{a \succ b}$ denote the set of all consumers who rank a over b. θ is called a general threshold (GT) for a recommendation set W if

$$a \in W \text{ whenever } \frac{|N_{a \succ b}|}{|N|} \geq \theta \text{ for all } b \in A \setminus \{a\}$$

A recommendation set is θ-consistent if θ is a GT for W.

The intuition here is that a low GT allows consumers to predictably push items "on the agenda" by gathering a *sufficiently large coalition* that rank those items high.

Majority Support (MS)
Note that the number of pairwise majority contests for a given news item is $|N| - 1$. Then σ is called a majority support (MS) threshold for a recommendation set W if

$$a \in W \text{ whenever } \frac{|\{b : \frac{|N_{a \succ b}|}{|N|} > \frac{1}{2}\}|}{|N| - 1} \geq \sigma \text{ for all } b \in A \setminus a$$

A recommendation set is σ-consistent if σ is a MS threshold for W.

In contrast to GT, a low MS threshold allows consumers to kick items off the agenda by establishing a *majority coalition* that ranks those items low.

Note that the last two properties are extensions of the Condorcet consistency axiom for the multiwinner case: Both axioms require that the Condorcet winner, if it exists, is in the winning set.

Gini-Coefficient (GINI)
Define the Gini-Coefficient (GINI) of a recommendation set as follows:

$$G(W) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |u_i - u_j|}{2nu(W)}$$

We care about GINI since unequal distributions of utility increase the likelihood that the worst-off consumers lose interest which would defeat the point of a common recommendation set.

4 Recommendation Rules

In this section, we define three extensions of k-multiwinner voting rules and introduce two new voting rules designed to perform optimally on the AUM and GT desiderata, respectively.

4.1 Extending Multiwinner Rules

We chose to adapt three k-multiwinner voting rules proposed by Elkind et al. (2017) for our setting: A budgeted Plurality rule (BuPLU) as a baseline, a budgeted Borda rule (BuBOR) as a more sophisticated representative of the positional scoring rules and a budgeted Copeland rule (BuCOP) to represent the Condorcet extensions. All of these rules assign a score to each item based on the current profile: $S : \mathcal{L}^n \times F \times A \to \mathbb{R}$; for brevity we will henceforth refer to all rules that assign a score by *fit-by-score rules* and denote an item's score, given a profile and a rule, as $S(a)$, instead of $S(\mathcal{R}, F, a)$.

The *fit-by-score rules* use their respective scores to recommend items in the following way: Start with the complete budget B and put the highest scoring budget-fitting items in the recommendation set. Then do the same for the remaining budget. Continue until the budget is filled. More formally, we define recursively:

$$W_1 = \{\arg\max_{a \in A} S(a)\} \cap \mathcal{W}_B$$

$$W_k = W_{k-1} \cup (\{\arg\max_{a \in A \setminus W_{k-1}} S(a)\} \cap \mathcal{W}_{B-C(W_{k-1})})$$

Then the recommended set is:

$$F(\mathcal{R}, C(A), B) = W_{k=|A|}$$

4.2 Rules Designed for Optimal Performance

It is noteworthy that *fit-by-score* combined with the Copeland-Score performs optimally with respect to MS. It elects the winning set with the lowest possible MS threshold σ. Intuitively, this means that it enables majority coalitions to kick an item off the agenda even if they dislike that item only a bit. Similarly, we designed rules to perform optimally with respect to AUM and GT.

Lowest General Budget-compatible Threshold (LGBT)

The first such rule is the *Lowest General Budget-compatible Threshold Rule* (LGBT). LGBT is designed to yield optimal results with respect to GT. To achieve this, we start by defining an item's θ-Score as follows:

$$\theta(a) = \min_{b \in A \setminus \{a\}} \frac{|N_{a \succ b}|}{|N|}$$

Then LGBT applies the *fit-by-score* method to elect the recommended set.

LGBT recommends the set that is optimal for minority preferences in the sense that it chooses the recommendation set with the lowest possible general threshold, thus enabling comparatively small coalitions to push "their" issues onto the agenda. For this it is sufficient but not necessary that a plurality of $\theta|N|$ voters rank the desired item first.

Budgeted Utility Optimization (BUM)

The second novel recommendation rule is the *Budgeted Utility Optimization* (BUM). BUM is designed to perform optimally with respect to AUM. It selects that set of articles which has the greatest overall utility for the consumers while fitting the budget.

$$F(\mathcal{R}, C(A), B) = \arg\max_{W \in \mathcal{W}_B} u(W)$$

Although computing the BUM recommendation is an NP-hard problem because it corresponds to the *0-1 Knapsack Problem*, there exist algorithms to solve it in pseudo-polynomial time. We used such an algorithm for our simulations.

5 Proofs

As a first step in assessing the performance of our rules, we obtained results regarding axiom satisfaction by way of proof. We only present proof sketches here.

Proposition 1. *All rules satisfy Non-Imposition*

Proof. Immediate. For any set within the budget limit consider elections with no other options.

Proposition 2. *All considered rules satisfy Consistency*

Proof. Omitted

Proposition 3. *All considered rules satisfy Homogeneity*

Proof. Consider any profile $\mathcal{R}$ and a budget B. Notice that in an election with t copies of $\mathcal{R}$ the hierarchy of Plurality- and Borda-Scores is not changed, as change in the scores is then proportional. Therefore, the winners of BuPLU and BuBOR rules are preserved. As utilities are captured in terms of Borda-Scores, the winner of BUM rule is also preserved. Also, the results of pairwise majority contests are preserved, as for any pair of items x, y the number of consumers who prefer x to y is multiplied by the same constant as those with opposite preferences. So LGBT and BuCOP rules preserve the winner. □

Proposition 4. *All considered rules satisfy Monotonicity*

Proof. We need to proof two claims presented in the Monitonicty axiom. We proceed by using defined $\mathcal{R}$, $\mathcal{R}'$ as in the axiom. For BUM, denote by $u_{\mathcal{R}}, u_{\mathcal{R}'}$ respectively the utility functions corresponding to $\mathcal{R}, \mathcal{R}'$. Let $a \in W$ and b be a's above neighbor in some consumer i's ranking.

For the first claim note that either $a, b \in W$ (I) or $b \notin W$ (II). For (II) the desired result is immediate. For (I), note that $u_{\mathcal{R}'}(\{a, b\}) = u_{\mathcal{R}}(\{a, b\})$. In addition if $b \in W'$ then a fortiori $a \in W'$. Hence suppose that $a, b \notin W'$. Then, since W' is utility maximizing, there must be some set of alternatives $X \subseteq A \setminus W$ s.t. $C(X) \leq C(\{a, b\})$ and $u_{\mathcal{R}'}(W \setminus \{a, b\} \cup X) > u_{\mathcal{R}'}(W)$. Equivalently

$$u_{\mathcal{R}'}(W) - u_{\mathcal{R}'}(\{a, b\}) + u_{\mathcal{R}'}(X) > u_{\mathcal{R}'}(W)$$

But none of these utilities have changed between $\mathcal{R}$ and $\mathcal{R}'$ and thus

$$u_{\mathcal{R}}(W) - u_{\mathcal{R}}(\{a, b\}) + u_{\mathcal{R}}(X) > u_{\mathcal{R}}(W)$$

But then W is not utility maximizing given B - a contradiction. We omit the proof for the latter claim.

For the fit-by-score rules: Denote by S, S' respectively the scoring functions corresponding to the profiles $\mathcal{R}, \mathcal{R}'$. (1) Since $S'(b) = S(b)$ for all $b \in A \setminus \{a, b\}$, $S'(a) \geq S'(b)$ for all b s.t. $C(b) \leq C(a)$. For any b s.t. $C(b) > C(a)$, either $S'(a) \geq S'(b)$ or $b \notin W'$ since it can not be fit into the budget by the fit and score method. But then $a \in W'$ as desired. We omit (2)

$\square$

Proposition 5. *No budgeted voting rule satisfies Committee-Monotonicity*

Proof. Consider a profile with only one consumer and two items. The only preference order is $(x_1 > x_2)$. Now suppose $C(x_1) = 9$, $C(x_2) = 1$ and $B = 8$. Then x_2 will be elected under any rule as only this item fits into the budget. However, if $B = 9$, clearly x_1 will be elected under any considered rule. So Committee-Monotonicity is always violated. $\square$

Proposition 6. *BUM and BP do not satisfy Unanimity*

Proof. For BUM, consider elections with budget $B = 10$, a set of items with corresponding costs $X = \{x_1 : 10, x_2 : 2, x_3 : 2, x_4 : 2, x_5 : 2\}$ and a profile of users with 2 consumers with the same preference orders $(x_1 > x_2 > x_3 > x_4 > x_5)$. Notice that following the unanimity axiom $\{x_1\}$ should be elected. But by summing up the (Borda-)utilities one can easily see that BUM elects $\{x_2, x_3, x_4, x_5\}$. For BuPOL, consider the same examples with but now $C(x_1) = 2$. Now x_1 is a plurality winner and the rest of the recommendation set will be filled by tie-breaking. But then x_5 might be recommended.

$\square$

Proposition 7. *All other rules satisfy Unanimity*

Proof. Consider a profile $\mathcal{R}$ such that there is a set of options W such that $W = B$ fits the budget maximally and W is ranked as a top set by all consumers. Then it is easy to see that any $x \in W$ has more Borda points than any $x' \notin W$, so all $x \in W$ are ranked strictly higher than any $x' \notin W$. So all $x \in W$ are elected as they can all fit in the budget together and no other option is elected due to maximality. So W is the winning set. For the BuCOP rule notice that all options in W win majority contests against all options outside of it. So, options in W have more Copeland points than other options, so W is elected. The argument for LGBT is symmetric. $\square$

6 Simulations

On top of proving results for the axioms, we investigated how the rules behave with respect to the metrics of performance outlined earlier.

6.1 Method

For this purpose, we first generated a large number of profiles and then automatically checked the performance of rule-profile pairs with respect to the four desirable properties GT, AUM, MS and GINI.

To account for different types of consumer populations, we created seven different profile-types, each representing a possible distribution of individual preferences over a population of 5000 consumers. A profile-type consisted of noisy copies of a number of *base preference*. We specified base preferences, copied them so as to get 5000 individual preferences and then applied noise to model variety amongst individual consumers. We employed a probabilistic model to swap items in the preferences, where farther swaps occurred with smaller probability than closer swaps. For all seven profile-types we generated 100 profiles each for the cases of 10 and 20 news items.

In terms of the results presented below, the most important profile-types are the following two with two base preferences each: For the first, the two base preferences are simply noisy copies of a third underlying preference order. We randomly generated one preference, applied high noise to it twice to get two base preferences and then generated 2500 moderately noisy copies of both. This resulted in profiles with two *clusters* of fairly similar consumer groups whose orderings agree mostly, but not completely, amongst themselves. We call them similar-cluster profiles. For the other profile-type with two base preferences we randomly generated a preference order, made 2500 noisy copies of it, then flipped it upside down and generated 2500 noisy copies of the reverse order. We call these polarized-cluster profiles.

So, we generated a large number of profiles with 5000 consumers each that vary across two dimensions. They are either made up of preference orders over 10 or 20 items, and the consumers in the profiles are grouped in different types of *clusters*, that is, closely scattered around an underlying base preference.

6.2 Results

We then studied how a) different profiles, all positioned somewhere on these two dimensions, affect performance on average with respect to our desired properties *for all of our rules* and b) how the rules perform against each other *for all profiles*. This allowed us to test for differences across distinct profiles for each rule, say, whether the number of candidates affects UM under the BuBOR, or whether highly polarized voter populations tend to drive up the Gini-coefficient more for BuCOP than for BUM. Our data analysis yielded a large number of results. We present here the most interesting. To test the significance of the obtained results, we employed ANOVA in combination with the Tukey post hoc analysis.

Results of Type a)

Firstly, concerning AUM, we found that for all profiles with 20 candidates and two clusters of consumers, the rules performed on average 5 to 10 percentage points worse on polarized-clusters populations than on similar-clusters populations.

Secondly, with respect to GT and profiles with two clusters of consumers, all rules perform between 30 and 55 percentage points worse on the polarized-cluster populations than on the similar-cluster populations. This can be seen in Figure. For LGBT, these differences were significant on a 95% confidence level for both the 10 and 20 items case. For BuBOR, BuCOP and BuPLU they were only significant for the profiles with 10 items.

Finally, comparing the mean GINI of all profiles with 20 items with the mean GINI of all profiles with 10 items yields that profiles with 20 candidates on average have a 15 and 22 percentage points higher GINI than profiles with 10 candidates. Except for BUM, all of these differences are significant on the 95% confidence level.

The a)-type results indicate that polarization of the readership leads to both lower utility and higher GT indicating what one might call a *cost of polarization*. This is bad news for hopes to find common recommendations for divided readerships. An avenue for future research could be the search for a recommendation rule that minimizes the cost of polarization.

We were somewhat puzzled by the increase in GINI correlated to an increase in the number of candidates. We attempted normalizing the utilities speculating that the much larger absolute difference in utility scores for 20-news-item-profiles vs. 10-news-item-profiles might cause the effect. However, the result remained robust. Thus, further research will be needed to get to the bottom of this.

Results of Type b)
We found that both BuBOR and BuPLU perform significantly worse with respect to AUM than BUM on the 95% confidence level. Perhaps surprisingly, while BuCOP and LGBT performed worse on AUM on average, the difference to BUM was not significant.

With respect to GT, we found that for both for the populations with 10 and 20 items, the LGBT rule performs significantly better than the other rules on the 99%-level. The difference is most pronounced of the BUM rule. For 10 items scores BUM on average 19.5 percentage points higher (worse) than LGBT. In the case of 20 items, BUM scores on average 21.8 percentage points higher (worse) than LGBT.

For the GINI we found that for 20 and 10 items, respectively, BUM performs 3.8 and 10.6 percentage points worse than LGBT, with significance on the 99% level. There were no significant differences between LGBT and the other score-and-fit rules.

Another type b) result pertains to MS. As expected, BuCOP performed best with respect to this axiom and significantly so compared to all other rules on the 99% interval. Less obviously, we also found that BB outperforms the remaining rules with a significance on the 99% level. Thus, both BUM and LGBT only did approximately as well as the baseline BuPLU rule with respect to MS.

We conclude that no single rule performs optimally with respect to all desirable properties and types of profiles. As expected we have a "three-way-tie" between BUM, LGBT and BC when it comes to AUM, GT and MS. However, looking at the other properties, a tendency can be established: First, it is notable that BUM is the only rule besides BP that fails the

Unanimity axiom. Secondly, while the BUM rule performs optimally on the AUM, the difference to BC and LGBT was not significant which indicates that BC and LGBT come close to the optimum. A further point to make is that while LGBT and BC are tied with respect to GINI, they clearly beat BUM in this respect.

To sum up: Although BUM maximizes utility by construction, this only leads to little added utility compared to BC and LGBT. Moreover, this additional utility comes at the cost of failing unanimity, markedly unequal utility distribution amongst the consumers, higher lowest general thresholds and higher lowest majority support thresholds. All of this arguably indicates that BUM is overall a worse rule then BC and LGBT.

7 Conclusion

In this paper we presented a formal Social Choice framework to recommend a common core of news items to a (possibly diverse) group of consumers given a budget constraint. For this we adjusted multiwinner axioms relevant to this setting and devised further performance measures which provide different ways to evaluate the recommended set. With these criteria in mind we introduced five rules, two of them novel, which were designed to perform especially well. For all of the rules we checked axiom satisfaction via proofs. Furthermore, we ran simulations in order to assess the rules against our performance measures. For these simulations we defined multiple population-types. For each type we generated multiple profiles. We then applied our voting rules to the profiles and performed statistical analysis on the results.

From our research we were able to draw two kinds of conclusions. Firstly, population structure greatly affects how well the rules perform. Notably, there is a *cost of polarization* leading to less desirable outcomes in polarized societies. Secondly, the BUM rule only achieves marginal improvement in utility at the expense of performance on other measures. Thus we conclude that LGBT and BC are the best rules in this setting known so far.

References

Elkind, E., Faliszewski, P., Skowron, P., & Slinko, A. (2017). Properties of multiwinner voting rules. *Social Choice and Welfare*, *48*(3), 599–632. Retrieved from `http://dx.doi.org/10.1007/s00355-017-1026-z` doi: 10.1007/s00355-017-1026-z

Lu, T., & Boutilier, C. (2011). Budgeted social choice: From consensus to personalized decision making. In *Ijcai* (Vol. 11, pp. 280–286).