Collective Recommendations

Haukur Jónsson
haukurpalljonsson@gmail.com
and Silvan Hungerbühler
silvan.hungerbuehler@bluewin.ch
and Grzegorz Lisowski
grzegorz.adam.lisowski@gmail.com
and Max Rapp
maxgrapp@gmail.com

Abstract

This is a placeholder for the glorious abstract yet to come.

1 Introduction

Media are in crisis. Between highly personalized social media - as Facebook or Twitter - that drive political fragmentation, partizanship and the absence of a common factual base in society on the one hand, and waning interest in the traditional, entirely unpersonalized newspaper on the other, there is little space to maneuver for news providers. How to balance the publics demand for news that is relevant to collective interests of both large and small groups in society and, at once, reflects individual interests.

The present paper takes a step towards solving this problem by designing and testing a number of recommendation mechanism 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 need not be of concern here; depending on the concrete application, the data can be thought of s 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 newspapers title page, for example, should not exceed its character limit. Likewise, there are relations between the essential articles and the rankings by the individuals one would like to see respected by a recommendation mechanism. By way of example, 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 what they recommend. We employ formal tools provided by *Social Choice Theory* to to analyze benefits and drawbacks of various possible ways to determine a set of essential news items - e.g. newspaper articles - for a group, given each member's individual preferences over the topics - eg. politics, sport, business. 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: Section 2 we provide the formal definition of the recommendation problem as we want to study it. In Section 3 we formally present the desirable properties a collection of recommended articles ought to have. In Section 4 we propose three rules for the task of turning individual preferences into recommendations. Section 5 contains the presentation and discussion of our simulation results as well as the theorems, while Section 6 concludes.

2 Formal Framework

This section specifies the formal framework we use. There is a set of *n*ews items $A = \{a_1, ..., a_m\}$, a subset of which are the *r*ecommended items $W \subseteq A$. Each item in A is assigned a specific *c*ost by the function $C : A \to \mathbf{R}$. Depending on the context, cost can be interpreted, for example, as the time it takes to read an article, the cognitive ressources it takes a consumer to digest it or simply character length. As these ressources are limited we assume *b*udget $B \in \mathbf{R}_{\geq 0}$. The set of *n*ews consumers $N = \{n_1, ..., n_n\}$ each have an indivual preference over A. This is represented by a complete order over the items $\mathcal{L}(A)$ for each consumer. For all consumers together this yields *p*rofiles of preferences $\mathbf{R} \in \mathcal{L}^n$.

The recommendation mechanism then is a function from profiles to recommended items: $\mathcal{L}^n \to W \subseteq A$. For some recommendation rules we assume a *v*alue function V which specifies how much a consumer values an option in her ballot amongst the recommended items, given her preferences. It takes as an input the consumer, the profile and an element of A, $V : \mathcal{L}^n \times N \times A \to \mathbf{R}$.

V would be akin to a Borda scoring rule, for example, in the concrete case where V outputs the value $m - 1$ for all consumers' candidates in the top position, $m - 2$ for the candidates in the second position and so forth.

Concerning *s*ocial value, then, there are two fundamental positions one could take. On the one hand, one a *u*tilitarian view we could measure social welfare as the sum of each all consumers' individual value. On the other, one could also adopt some *e*galitarian point of view which takes the differences between consumers' values into account.

3 Desirables & Dimensions of Performance

The following are a bunch of things we would like a recommendation mechanism to do. We will describe these desiderata formally and describe their relation and relevance to our goal stated in the introduction.

3.1 Desirables

Regret Minimization

Let $\mathcal{W}_\mathcal{B}$ be the set of all budget-compatible elements of $\mathcal{P}(A)$. Regret is minimized (or utility is maximized) if:

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

θ-Minority Consistency

Let a : b denote the majority contest between alternatives a and b and $\#_a(a : b) \in [0, 1]$ the share of the vote that a wins against b in that contest. A recommendation set is θ-minority consistent if

$$a \in W \text{ whenever } \#_a(a : b) \geq \theta \text{ for all } b \in A \setminus \{b\}$$

δ-Equality

Define the Gini-coefficient of a recommendation set as follows:

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

For $\delta \in [0, 1]$, a recommendation set is δ-egalitarian if $G(W) \leq \delta$.

Cost Distribution

There are two polar extremes here: On one end one could conceive a rule that picks chunky, expensive items (long, in-depth articles); on the other end rules that pick snacky, cheap items (Vice). In principle, any distribution over article prices is conceivable.

A bunch of more rules we yet have to get from the article.

4 Recommendation Rules

Here come a bunch of voting rules that we think will perform well with respect to above desiderata.

4.1 K-plurality rule

The *k-plurality rule* with $k < |A|$ is a *positional scoring rule* with the same scoring vector as the normal *plurality rule*, $(1, 0, \ldots, 0)$, but instead of electing the alternative(s) with the highest score it elects the alternative(s) with the highest score, if the number of winners is strictly less than k then the alternative(s) with the second highest score is elected. This is done until $|W| \geq k$. If $|W| > k$ then a tie-breaker should be applied on the last iteration.

Computationally feasible.

4.2 Schulze rule

Has many nice properties and is closely related to the 20% minimal requirement.

Computationally feasible.

4.3 Extended Θ-Smith

The Θ-Smith set is the smallest non-empty set of candidates s.t. each member of the set defeats every other member outside the set in Θ% of cases.

The Θ-Smith set is a Condorcet extension.

A suggestion of an algorithm which selects a "good" Θ-Smith set. Compute the Smith set for $\Theta = 50$ (the normal Smith set), check if it satisfies the cost restricitons. Second, check the upper-bound by finding the Smith set of $\Theta = 100$, if it satisfies the cost restriction, select it. Depending on the case the cost restriction was not satsfied we need to iterate over the interval in which the cost restriction break, start iterating from $\Theta = 50$ and increase/decrease Θ by $(100 - 50)/2 = 25$ or $(50 - 0)/2 = 25$ until a set is found which satisfies the cost restrictions (when found, there might be multiple.).

Computationally feasible.

4.4 θ-rule

Instead of a Condorcet Committee, we could elect the *set of θ-winners* Θ, that is the set of all alternatives that win at least θ% in all majority contests. Given the budget constraint, we choose the smallest θ that is still compatible with the budget.

We define Θ as follows: let $a : b$ denote the majority contest between a and b and $\#_a(a : b) \in [0, 1]$ the share of the vote that a wins against b in that contest. Then $\Theta = \{a \in A : \text{ for all } b \in A, \#_a(a : b) \geq \theta\}$. Then W=$\Theta$ subject to the following optimization problem:

$$\min \theta \text{ subject to } \sum_{a_i \in W} c(a_i) \leq B$$

4.5 Utility Optimization

We could also cast the problem of coming up with the best recommendation, given the items' costs as well as the budget, as a maximization problem. What we try to maximize is the sum of all the consumers' values by choosing W, where a consumer's value of seeing some item in W is represented by a vector $\mathbf{s}$. For each consumer we only count the value from the news items that are actually in the recommended set:

$$\max_W \sum_{j=1}^{n} \sum_{i=1}^{m} \mathbb{1}[a_i \in W] V(\mathbf{R}, n_j, a_i)$$

Of course, Equation 1 is trivially solved by $W = A$. But the interest in solving it comes from adding the budget constraint.

This problem, however, is the well-studied *0-1 Knapsack* problem. Given a set of items, each with a weight and a value, what is the most valuable knapsack that you can put together without exceeding its maximal load. In our case the weight corresponds to an article's cost, the value to the sum of all consumers' value if the item is included in W, and the maximal load corresponds to our budget B.

Although the optimization problem is NP-hard, there exists pseudo-polynomial algorithms using dynamic programming as well as polynomial-time approximation schemes.

5 Simulations

How this section is structured, what it contains.

5.1 Method

How we proceeded in empirical tests.

5.2 Results

What results we obtained in empirical tests.

5.3 Discussion

How we interpret the empirical tests

5.4 Proofs

What proofs we have. Details can be given in appendix.