Fuzzy D'Hondt's Algorithm for On-line Recommendations Aggregation

Ladislav Peska　　　　　　　　　　　　　　　　　　　　　PESKA@KSI.MFF.CUNI.CZ
Dept. of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague

Stepan Balcar　　　　　　　　　　　　　　　　　　　　　BALCAR@KSI.MFF.CUNI.CZ
Dept. of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague

Editors: João Vinagre, Alípio Mário Jorge, Albert Bifet and Marie Al-Ghossein

Abstract

In this paper, we present Fuzzy D'Hondt's algorithm suitable to aggregate lists of recommended objects originating from various base recommending methods. The algorithm is inspired by D'Hondt's election method used to a proportional conversion of votes to mandates in public elections. We enhance the original approach to enable fuzzy candidate-party membership, propose a gradient learning of per-party votes assignments and utilize it for iterative on-line aggregation of recommendations. Main features of the proposed algorithm are ability to iteratively learn relevance of individual base recommenders (parties), ability to account for multiple item's memberships and capability to provide proportional representation of base recommenders w.r.t. their results as well as fair ordering of the final list of recommended items. Fuzzy D'Hondt's aggregation method was evaluated in on-line A/B testing against state-of-the-art approach based on multi-armed bandits with Thompson sampling and achieved competitive results.

Keywords: Fuzzy D'Hondt's algorithm, aggregation of recommendations, A/B testing

1. Introduction

Recommender systems (RS) belong to the class of automated content-processing tools, aiming to provide users with unknown, surprising, yet relevant objects without the necessity to explicitly query for them. The core of recommender systems are machine learning algorithms applied on the matrix of user to object preferences. As such, recommender systems are highly studied research topic as well as extensively used in real-world applications.

Over the decades of the RS development, a wide range of algorithm types were proposed, e.g., collaborative, content-based, hybrid, session-based, graph-based or context-aware to name just a few. Obviously, also the members of the same group of algorithms may differ significantly from each other. Furthermore, the availability of background information (e.g., various attributes in content-based RS) affects the results of otherwise identical algorithms as well. To sum up, there is a very large space of available RS algorithms and its variants and it is often unclear, which algorithm(s) to employ for a particular recommending task.

A common practice in such situations is to employ on-line evaluation of algorithms via A/B testing, or some regret-minimizing methods, such as selection of algorithms via multi-armed bandits (Audibert et al., 2009). However, in some scenarios, it may be important to go beyond a simple variant selection. In numerous works, e.g. Di Noia et al. (2017),

the diversity was identified as one of the key features of the list of recommendations. In addition to relevance and diversity, Steck (2018) examined the problem of fairness in RS, i.e., proportional representation of all user's interest in the list of recommendations. Common aggregation techniques, such as weighted average, tend to over-sample the major class, if there is only a small intersection in the individual lists of recommendations. Similarly, most of the contemporary individual RS techniques would over-sample items corresponding to the major user's interest and ignore his/her other topics.

Both the diversity of recommendations and the fair representation of user's topics may be achieved by the utilization of an ensemble of base RS algorithms[1] and some suitable aggregation of their predictions. Nevertheless, each base RS as well as each recommended item is linked with different relevance or probability of success, which should be attributed in the aggregation process.

We observed that the above mentioned circumstances are highly similar to the process of public elections. During elections, political parties propose an ordered list of candidates, where the ordering can be understood as a relevance of a candidate to the party. During elections, each party receives a certain portion of votes, which can be considered as party's relevance. Votes are further transformed into mandates in the considered assembly. The key part of the process is the transformation of votes to mandates, for which, several algorithms were proposed. In this paper, we focus on one of these algorithms, D'Hondt's election algorithm (D'Hondt, 1882), which is currently utilized in the election process of over 50 countries. D'Hondt's election algorithm focuses on proportional representation of each party through the mechanism of iterative decay of accountable votes upon the selection of a party representative. As such, one of the algorithm's feature is also a fair, deterministic ordering of the elected candidates.

There are three key differences between the recommendations aggregation and the public elections scenarios. First, RS algorithms usually provide more refined item relevance scores beyond a simple ranking. Second, the same candidate item may be recommended by multiple algorithms.[2] Finally, the volume of votes (i.e., algorithm's relevance) needs to be set by the system. Therefore, in order to address the fair recommendations aggregation challenge, we introduced several modifications to the original algorithm and proposed it as Fuzzy D'Hondt's aggregation algorithm with iterative voting updates. To the best of our knowledge, this is the first work aiming to utilize and evaluate D'Hondt's election method in the domain of recommender systems.

Utilization of multi-armed bandits in recommendations aggregation is closely related to our approach. We can specifically mention the BEER framework by Brodén et al. (2018). However, in BEER, authors did not consider the possible synergic effect, where a single object is recommended by multiple RS. Furthermore, BEER is based on Thompson Sampling, which does not focus on proportional representation, but performs a stochastic selection of the best method w.r.t. current level of knowledge. Similarly to our work, Steck (2018) aims on proportional representation of topics, however he focused on re-ranking of a single list of recommendations instead of RS aggregations. To sum up, main contributions of the paper are as follows:

---

1. For instance, each individual RS may exploit different recommending paradigm (and therefore provide diverse recommendations in general) or focus on individual user's topic separately.

2. Using the original metaphor, we may say that the party-candidate membership is fuzzy.

- Extension of original D'Hondt's election algorithm to handle fuzzy affiliation of candidates to parties.

- Proposal of an on-line iterative votes assignment algorithm based on positive and negative implicit feedback.

- Evaluation of the proposed Fuzzy D'Hondt's algorithm on the recommendations aggregation task in an on-line A/B testing.

2. Methods

2.1. D'Hondt's Election Algorithm

In order to make this paper self-contained, let us first describe the original D'Hondt's Election Algorithm.[3] Suppose that we have a list of political parties $\mathcal{P}$. For the process of elections, each party $p_i \in \mathcal{P}$ creates an ordered list of candidates $C_i = [(k, c_k^i)]$, where k and c_k^i denotes a rank and a candidate proposed by party p_i on k-th position. During the elections, some amount of votes v_i is assigned to each party p_i. D'Hondt's is initialized with per-party accountable votes a_i equal to the original volume of votes v_i. Then, the algorithm iterates over the following steps until the demanded volume of mandates is met.

- Select party p_{best} with the highest volume of accountable votes, $best = \mathrm{argmax}_{\forall i}\, a_i$.

- Select best ranked candidate (k, c_k^{best}) of party p_{best}, which was not selected before and append c_k^{best} to the list of representatives.

- Decrease the volume of accountable votes of party p_{best} as follows: $a_{best} = v_{best}/(k+1)$.

There are several notable features of the presented method. First, D'Hondt's method solves the problem of proportional transformation of votes to mandates through the accountable votes reduction upon candidate selection. This is by design a greedy solution iteratively selecting the current best method (party) and thereafter reducing its preference. Based on the form of the votes reduction, D'Hondt's algorithm slightly favours large parties (i.e., highly preferred methods) over scattered small parties (Schuster et al., 2003). This seems suitable also in the case of recommendations aggregations, where methods with negligible preference should be ignored as a noise (note that even random recommendations usually attract some users). Second, the method always selects the most preferred one out of the remaining party's candidates and therefore, it iteratively constructs the ordered list of current best candidates.

2.2. Fuzzy D'Hondt's Aggregations

The original D'Hondt's method reflects well the need for proportionality while aggregating results of multiple base recommenders and it copes well with the problem of highly diverse set of recommended objects. Nonetheless, it ignores the possible synergic effect, if the same

---

3. For the sake of simplicity, we do not describe some of D'Hondt's method variants. See, e.g., Gallagher (1991) for more details.

Algorithm 1 Fuzzy D'Hondt's Aggregation - Mandates Selection

Input: set of parties $p_i \in \mathcal{P}$, set of candidates $c_j \in \mathcal{C}$, relevance matrix $\mathbf{R}$, per-party votes $v_i \in V$, volume of mandates m
Output: ordered list of selected candidates C_s

$\forall i : a_i = v_i, k_i = 1$
$C_s = \emptyset$

for $l \in [0, ..., m]$ **do**
    select c_{best} according to Eq. (1)
    append c_{best} to C_s
    remove c_{best} from the set of candidates $\mathcal{C}$
    $\forall i : k_i = k_i + r_{i,best}; a_i = v_i/k_i$
    # increase the per-party counters of selected candidates and update accountable votes

end for
return C_s

---

candidate item is recommended by multiple base recommenders. In order to address also this issue, we propose a fuzzy extension of original D'Hondt's algorithm as follows.

Suppose that $\mathcal{C}$ is a set of all candidates proposed in the elections across all parties. Furthermore, suppose that instead of a simple ranking of candidates, each party can express the relevance of its candidates as follows: $C_i = [(\hat{r}_k^i, c_k)]; \hat{r}_k^i \in (0, 1], c_k \in \mathcal{C}$. We can naturally extend this per-party rating to the full list of candidates by assuming zero relevance: $\forall c_j \in \mathcal{C}, c_j \notin C_i : \hat{r}_j^i = 0$. In this way, we can construct a matrix of per-party ratings of all candidates $\mathbf{R}$, such that each cell $r_{i,j} \in \mathbf{R}$ contains the relevance of j-th candidate to the i-th party.

Similarly to the original method, some amount of votes v_i is assigned to each party p_i and the algorithm iteratively selects candidates until the desired volume of mandates is reached. However, instead of selecting the best party and then its best remaining candidate, Fuzzy D'Hondt's method selects candidate with the highest overall relevance, which is calculated as a weighted sum of accountable votes, where the weights are per-party relevance scores of candidates.

$$c_{best} = \underset{\forall c_j \in \mathcal{C}}{\operatorname{argmax}} \Big(\sum_{\forall p_i \in \mathcal{P}} a_i * r_{i,j} \Big) \tag{1}$$

After the selection of the best actual candidate, the accountable votes of parties are decreased in proportion to the per-party candidate's relevance. Further details can be found in Algorithm 1. In our use-case, we represent recommenders as parties and top-k recommended items correspond with the selected candidates. Candidates' relevance scores w.r.t. each party are normalized into a unit vector to prevent bias.

It is easy to see that the proposed fuzzy extension of D'Hondt's method provides the same results as the original method, if the lists of candidates are disjoint and the relevance score for a candidate on k-th position is defined as $1 - k * \epsilon$ for some sufficiently small ϵ. The crucial part of the algorithm proposal is the proportional reduction of accountable votes

according to the fuzzy per-party affiliations, which prevents over-sampling of candidates preferred by a single method.

2.3. Iterative Votes Assignment Algorithm

Both D'Hondt's election method and it's proposed fuzzy extension suppose that votes are supplied by a third party.

We propose to assign per-party votes dynamically, in accordance with the observed negative and positive implicit feedback. In the evaluation, we assume that clicking on a recommended item is an expression of positive preference (of the respective item), while ignoring recommendations is an expression of (weak) negative preference. Nonetheless, the actual selection of feedback events may differ according to the considered domain. Proposed algorithm considers per-party votes v_i as variables and aims to maximize the weighted sum of votes for preferred candidates $c_j \in \mathcal{F}^+$ and minimize the same criterion for ignored candidates $c_j \in \mathcal{F}^-$ at the same time (Eq. 2).

$$\max_{\forall v_i} \Big(\sum_{\forall c_j \in \mathcal{F}^+} \big(\sum_{\forall p_i \in \mathcal{P}} v_i * r_{i,j} \big) - \lambda_{neg} \sum_{\forall c_j \in \mathcal{F}^-} \big(\sum_{\forall p_i \in \mathcal{P}} v_i * r_{i,j} \big) \Big) \tag{2}$$

Similarly as Frigó et al. (2017), we opted for on-line learning approach that performs a single stochastic gradient ascend (descend) step for each positive (negative) feedback event. The update steps are

$$v_i = v_i + \eta_{pos} * \big(r_{i,j} - \sum_{\forall k \neq i} r_{k,j} \big) \text{ for positive feedback events}$$

$$\text{and } v_i = v_i - \eta_{neg} * \big(r_{i,j} - \sum_{\forall k \neq i} r_{k,j} \big) \text{ for negative feedback events,} \tag{3}$$

where η_{pos} and η_{neg} are learning rate hyperparameters. One advantage of such setting is the ability to adapt to performance changes of individual base recommenders over time. In order to prevent divergence of the votes assignment, we impose minimal and maximal boundaries for the volume of votes and linearly scale them so that the sum of all votes equals one. The method can be initialized either with uniform votes distribution, or based on the previous performance evaluation of base recommenders.

2.4. Design Choices, Remarks

Let us summarize some design choices and alternatives, which can be explored in the future. First, candidates selection process is deterministic by design (which is obvious, given the original task), but a stochastic implementation would be also possible, subjected to some minor changes in votes and/or candidate's preference definitions. While designing votes aggregation (Eq. 1), we opted for a simple solution, as this is a pioneering work of this research direction. However, other solutions, such as some S-norm aggregations are possible as well (Klement et al., 2000). The volume of results provided by base RS can be considered as an implicit hyperparameter of the method, where larger volumes generally favor items that are not optimal for any individual RS, but good-enough for most of them, i.e., with more recommended items, we are moving from *max-like* to *mean-like* aggregations. Finally,

utilization of weak negative feedback can be considered as a kind of regularization for votes assignment, preventing some early random successes to permanently bias the assignment.

3. Evaluation

In this section, we describe the evaluation of proposed Fuzzy D'Hondt's Aggregation method in on-line A/B testing against a state-of-the-art RS aggregation algorithm as well as some individual base recommenders. Let us first describe the dataset and base recommenders.

3.1. Dataset and Base Recommender Systems

Experiments were conducted on SLAN tour - a medium-sized Czech travel agency. The agency sells tours of various types to several dozens of countries. Some tours (such as trips to major sport events) are one-time only events, others, e.g., seaside holidays or sightseeing tours are offered on a similar schedule with only minimal changes for several years. All tours contain a textual description accompanied with a range of content-based (CB) attributes, e.g., tour type, meal plan, type of accommodation, length of stay, prices, destination country, points of interest etc. Agency's website contains attribute and keyword search GUI as well as some browsing and sorting options. Recommendations are displayed throughout the website, specifically on main page, browsed categories, search results and tour details. However, due to the importance of other GUI elements, recommendations are usually placed below the initially visible content.

In accordance with Kaminskas et al. (2017), we focused on item-to-item recommending models and utilized three base RS algorithms corresponding with three principal sources of data: object's CB attributes, their textual description and the history of users' visits. Expected output of a RS based on CB attributes are objects similar to the ones in a query. Stream of user's visits (i.e., collaborative filtering) reveals related, not necessarily similar objects. Text-based approaches also provide similar objects, however based on unstructured knowledge hidden in plain-text descriptions.

For each type of source information, we utilized one state-of-the-art algorithm: Skip-gram **word2vec** model (Mikolov et al., 2013) applied on the stream of objects visited by a user, **Doc2vec** model (Le and Mikolov, 2014) applied on the textual description of objects and finally, **cosine** similarity utilized on **CB** attributes.[4] Given a query of a single object, the outcome of base recommenders would be a list of top-k objects most similar to the query (w.r.t. cosine similarity on the considered data representation). We then derived numerous variants of the base algorithms, utilizing different hyperparameter settings (e.g., window and embedding sizes), several variants of user's history aggregation and options to enhance novelty or diversity of the resulting list of recommendations.[5] Word2vec algorithms were trained on more than one year of historical usage data (sequences of visited objects per user), while doc2vec and cosine CB models received description of all tours available within this period.

---

4. As a pre-processing, nominal attributes were binarized and numeric attributes were standardized.

5. Details can be found in Peska and Vojtas (2018)

7

3.2. Evaluation Protocol

In evaluation, we focused on on-line performance of algorithms through A/B testing. Evaluation was conducted on the travel agency's production server between July 6 and July 29, 2019 and each user was assigned one of the evaluated recommenders based on his/her ID. Proposed Fuzzy D'Hondt's algorithm was compared against two base RS: a word2vec model with diversity enhancements and a cosine CB. Those were the best performing RS in our previous work. In addition, we employed a recommendation aggregation technique based on multi-armed bandits, BEER(TS,SB) by Brodén et al. (2018). Both Fuzzy D'Hondt's and BEER(TS,SB) were supplied with the top-100 predictions of four base recommenders.[6] Both methods were initialized according to the past individual performance of base RS. Furthermore, we set η_{pos} and η_{neg} to 0.03 and 0.00012 respectively. We consider η_{pos} value to be a reasonable compromise between quickly converging and noise-prone systems, while η_{neg} was selected to maintain approximately half of the importance of the positive feedback (after the incorporation of positive vs. negative events ratio). Nonetheless, we plan to run a proper hyperparameter tuning in the future.

Based on the collected data, we evaluated two metrics: click through rate (CTR) and visit after recommend rate (VRR). CTR is a fraction between the volume of clicked and recommended objects and indicates that a recommendation was relevant for the user and caught his/her attention. VRR is a weaker criterion checking that after the object was recommended, the user also visited it (i.e., (s)he might not pay attention to recommendations, but the recommended object was probably relevant). Although VRR is weaker than CTR, we selected it as a main evaluation metric due to the higher volume of recorded events and also because recommended objects were often placed outside of the initially visible area and therefore CTR may underestimate the true utility of recommendations.

We further evaluated two derived metrics: CTR_U and VRR_U, where clicks (visits) are divided by the volume of unique users instead of the volume of recommended objects. We introduce these criteria, because applied RS may also have some impact on the length of user sessions, i.e., the volume of provided recommendations.

3.3. Results and Discussion

A total of 78356 objects were recommended to the 2299 users. The total volume of click-through events was 576 and the total volume of visits after recommendation was 7663. Table 1 contains overall results of the evaluation. We may observe that although Fuzzy D'Hondt's was outperformed by BEER(TS,SB) w.r.t. CTR, the results were opposite according to VRR. We hypothesize that this is an effect of stochastic sampling utilized by BEER(TS,SB), which results in more diverse recommendations through time and therefore may repeatedly attract user's attention. Therefore, we plan to evaluate a variant of Fuzzy D'Hondt's method with stochastic sampling or introduce relevance penalization through time for displayed items. In 3 out of 4 evaluated metrics, one of the aggregated methods outperformed both base recommenders.

---

6. Both base RS from A/B testing and two additional RS selected w.r.t. diversity of algorithms and their past performance: a doc2vec model and another word2vec model without diversity enhancements. For detailed parameters, see Table 2 (lines 3, 7, 10 and 12) in Peska and Vojtas (2018).
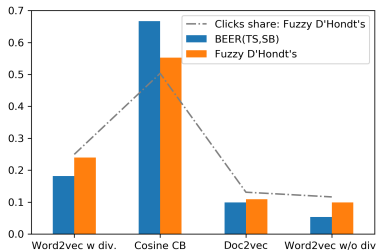
Figure 1: Utilization of base RS.

Algorithm	CTR	VRR	CTR_U	VRR_U
Cosine CB	**.0086**	.089	.262	2.707
Word2vec w. div.	.0071	.104	.220	3.250
BEER(TS,SB)	.0077	.078	**.282**	2.861
Fuzzy D'Hondt's	.0063	**.118**	.239	**4.504**

Table 1: Results of A/B testing.

In order to compare the models of aggregation methods, we exploited the log of recommended items. Figure 1 depicts the share of individual base recommenders on the final recommendations.[7] Both BEER(TS,SB) and Fuzzy D'Hondt's most preferred Cosine CB, which corresponds to the CTR results of base RS. Nonetheless, Fuzzy D'Hondt's algorithm provides more uniform representation of base RS, which, given the small differences of base RS's results, seems to be a legitimate choice. Also, the relative share of base RS on click-through events corresponds with their representation within recommended objects for Fuzzy D'Hondt's algorithm.

In general, we may conclude that if we are aiming on proportional representation of base RS,[8] Fuzzy D'Hondt's method seems to be a better choice than BEER(TS,SB).

4. Conclusions and Future Work

In this paper, we presented Fuzzy D'Hondt's algorithm for recommendations aggregation. It aims to achieve a proportional representation of the base RS according to their performance and to provide a fair ordering of items according to their overall relevance w.r.t. all base RS. In evaluation, we showed that the proposed algorithm provides more proportional distribution w.r.t. base recommenders and outperforms its competitors in three out of four evaluated metrics.

In this short paper, we only evaluated a single variant of the proposed algorithm on a single dataset, so in the future work, we plan to extend the evaluation part with additional domains and base RS, employ stochastic items sampling as well as the concept of relevance penalization through time for displayed items. Furthermore, we would like to explore also other election algorithms and their potential applicability to RS aggregation task.

Acknowledgments

---

7. In Fuzzy D'Hondt's, we considered the relative preference of object by each base RS (i.e., $v_i * r_{i,j}$).

8. E.g., because each base algorithm focuses on a different user's topic or recommending paradigm.

References

Jean-Yves Audibert, Remi Munos, and Csaba Szepesvari. Exploration-exploitation tradeoff using variance estimates in multi–armed bandits. *Theoretical Computer Science*, 410(19): 1876 – 1902, 2009. ISSN 0304-3975. doi: https://doi.org/10.1016/j.tcs.2009.01.016. URL http://www.sciencedirect.com/science/article/pii/S030439750900067X. Algorithmic Learning Theory.

Björn Brodén, Mikael Hammar, Bengt J. Nilsson, and Dimitris Paraschakis. Ensemble recommendations via thompson sampling: An experimental study within e-commerce. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pages 19–29, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-4945-1. doi: 10.1145/3172944.3172967. URL http://doi.acm.org/10.1145/3172944.3172967.

Victor D'Hondt. *Système pratique et raisonné de représentation proportionnelle*. C. Muquardt, 1882.

Tommaso Di Noia, Jessica Rosati, Paolo Tomeo, and Eugenio Di Sciascio. Adaptive multi-attribute diversity for recommender systems. *Inf. Sci.*, 382(C):234–253, March 2017. ISSN 0020-0255. doi: 10.1016/j.ins.2016.11.015. URL https://doi.org/10.1016/j.ins.2016.11.015.

Erzsébet Frigó, Róbert Pálovics, Domokos Kelen, Levente Kocsis, and András Benczúr. Alpenglow: Open source recommender framework with time-aware learning and evaluation. In *2017 Poster Track of the 11th ACM Conference on Recommender Systems, Poster-Recsys 2017*, pages 1–2. CEUR-WS.org, 2017. URL http://eprints.sztaki.hu/9324/.

Michael Gallagher. Proportionality, disproportionality and electoral systems. *Electoral Studies*, 10(1):33 – 51, 1991. ISSN 0261-3794.

Marius Kaminskas, Derek Bridge, Franclin Foping, and Donogh Roche. Product-seeded and basket-seeded recommendations for small-scale retailers. *Journal on Data Semantics*, 6 (1):3–14, Mar 2017. ISSN 1861-2040. doi: 10.1007/s13740-016-0058-3. URL https://doi.org/10.1007/s13740-016-0058-3.

Erich Peter Klement, Radko Mesiar, and Endre Pap. *Families of t-norms*, pages 101–119. Springer Netherlands, Dordrecht, 2000. ISBN 978-94-015-9540-7. doi: 10.1007/978-94-015-9540-7_4. URL https://doi.org/10.1007/978-94-015-9540-7_4.

Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, 2014. URL http://arxiv.org/abs/1405.4053.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *NIPS '13*, pages 3111–3119, USA, 2013. Curran Associates Inc. URL http://dl.acm.org/citation.cfm?id=2999792.2999959.

L. Peska and P. Vojtas. Off-line vs. On-line Evaluation of Recommender Systems in Small E-commerce. *ArXiv e-prints*, September 2018.

Karsten Schuster, Friedrich Pukelsheim, Mathias Drton, and Norman R. Draper. Seat biases of apportionment methods for proportional representation. *Electoral Studies*, 22(4):651 – 676, 2003. ISSN 0261-3794. doi: https://doi.org/10.1016/S0261-3794(02)00027-6. URL http://www.sciencedirect.com/science/article/pii/S0261379402000276.

Harald Steck. Calibrated recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 154–162, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5901-6. doi: 10.1145/3240323.3240372. URL http://doi.acm.org/10.1145/3240323.3240372.