



CrEx-Wisdom Framework for Fusion of Crowd and Experts in Crowd Voting Environment – Machine Learning Approach

Ana Kovacevic^{1,2}(✉) , Milan Vukicevic¹ , Sandro Radovanovic¹ ,
and Boris Delibasic¹ 

¹ Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia
ak20195017@student.fon.bg.ac.rs,
{milan.vukicevic, sandro.radovanovic,
boris.delibasic}@fon.bg.ac.rs

² Saga LTD, Belgrade, Serbia

Abstract. In recent years crowd-voting and crowd-sourcing systems are attracting increased attention in research and industry. As a part of computational social choice (COMSOC) crowd-voting and crowd-sourcing address important societal problems (e.g. participatory budgeting), but also many industry problems (e.g. sentiment analyses, data labeling, ranking and selection, etc.). Consequently, decisions that are based on aggregation of crowd votes do not guarantee high-quality results. Even more, in many cases majority of crowd voters may not be satisfied with final decisions if votes have high heterogeneity. On the other side in many crowd voting problems and settings it is possible to acquire and formalize knowledge and/or opinions from domain experts. Integration of expert knowledge and “Wisdom of crowd” should lead to high-quality decisions that satisfy crowd opinion. In this research, we address the problem of integration of experts domain knowledge with “Wisdom of crowds” by proposing machine learning based framework that enables ranking and selection of alternatives as well as quantification of quality of crowd votes. This framework enables weighting of crowd votes with respect to expert knowledge and procedures for modeling trade-off between crowd and experts satisfaction with final decisions (ranking or selection).

Keywords: Crowd voting · Experts · Machine learning · Clustering · Matrix factorization

1 Introduction

Inclusion of crowd in decision-making processes may not only result in greater crowd satisfaction, but also higher quality and timeliness of decisions even if they are compared to decisions made by limited number of experts [5]. This is due to phenomena of “Wisdom of Crowd” or “Collective Intelligence” that have theoretical roots in Condorsets Jury theorem. This theorem states that given a group of independent voters (a “jury”) that

have probability of correct outcome (alternative) $1 \geq p \geq 0$ and incorrect outcome of $1-p$, probability of choosing correct outcome by majority voting increases by adding more voters if probability of the correct outcome of each voter is greater than random choice (e.g. $p > 0.5$ in case of binary outcome). Even though this theory has strong assumption on voter independence that is not fulfilled in many real-world scenarios and several other limitations, it showed cutting edge results in many application areas and problems of ranking, selection, prediction, etc. Over the last few years, Collective Intelligence (CI) platforms have become a vital resource for learning, problem-solving, decision-making, and predictions [8] and led to development of numerous frameworks. Adequate technology support and desirable properties of crowd-voting systems led to wide acceptance of crowd/voting as a tool for solving both industry and societal problems [22].

In societal problems inclusion of crowd in decision making should lead to greater satisfaction and welfare. Additionally, “Wisdom of crowds” may be exploited in order to make high-quality decisions, while satisfying crowd opinion. Collection of votes from the general population can be encouraged by the following main reasons [22]: democratic participating in political elections and policymaking (e.g., law regulation [18]); solving issues from common interest (e.g., budget allocation – Knapsack voting and participatory budgeting [19] or resolving a different kind of issues in the field of education, health, etc.).

For many industry problems, companies adopt “Crowd Intelligence” in order to automate processes increase quality of their products and services and reduce costs. For example: choosing innovative ideas that should be adopted [20]; giving feedback on creative works [10]; making recommendations based on users’ critical rating [21]; stock market predictions [3]; selecting winners in competitions (e.g. TV music competitions such as Eurovision Song Contest, American Idol, etc.), and others.

However, implementation of knowledge and patterns identified in information collected from crowd in both societal and industry settings poses a significant challenge. Most of the problems that are inherited from assumptions of Condorcet’s jury theorem that are not fulfilled in most of the real world application problems. Some of the major problems are:

- Incompetence, lack of interest, favoritism and manipulation of the crowd for problem at hand [9],
- Bias in ordinal voting systems [6],
- Sparse and imbalanced data generated from crowd votes [2],
- Etc.

We hypothesize that exploitation of expert knowledge (even with single or limited number of experts) may address many problems of crowd voting quality, while preserving advantages of “Wisdom of Crowd” and “Collective Intelligence”.

In this paper, we present a framework that enables fusion of experts’ domain knowledge based on unsupervised machine learning approach. The main idea of the framework is to use limited number (or single) of expert inputs in order to weight crowd votes. In this way, we pose the problem of vote aggregation as a minmax problem: minimization of distance from experts and maximization of crowd satisfaction. We address this problem

by estimation of density and similarity of votes between crowd and experts through clustering and outlier detection. Additionally, we address the problem of sparseness of votes by using matrix factorization techniques that showed cutting edge results in the area of recommender systems based on collaborative filtering. Such factorization enables not only dimensionality reduction and solving sparsity problem, but also extraction of latent features that represent affinities of crowd and expert voters. Affinities in dense format enable definition of good quality distance/similarity measures but also estimation of voters' preferences towards alternatives that they did not voted for (or gave rating). In experimental part of this paper, we show usefulness of our approach on the Eurosong contest ranking problem. We compare the results in terms of both expert and crowd satisfaction by final ranks with two benchmarks: official Eurosong voting aggregation procedure and newly weighted voting procedure that does not exploit benefits of latent feature space.

The contribution of this paper is twofold:

1. We propose a framework for unsupervised machine learning based aggregation of crowd and expert opinions.
2. We provide an experimental evaluation of the framework and make additional insights on crowd performance based on characteristics of crowd and experts opinions.

2 State-of-the-Art

Exhaustive and systematic review on Collective Intelligence (CI) platforms including 9,418 scholarly articles published since 2000 recently is presented in [8]. Additionally, in our previous work [22] we provided detailed review and analyses of advantages and disadvantages of expert-based and crowd-based decision making systems that are summarized in Table 1. Thus, in this literature review we will focus only on research that is closest to current research with special focus on similarities and differences and compatibility between similar approaches and the one proposed in this paper.

Usage of matrix factorization in CI is not a new idea. There are a numerous examples where latent features are extracted to help the process of decision-making. One such example is filling missing values in crowd judgments [2]. Majority of the voters in the CI process express their judgments for only several alternatives (out of a much larger set of alternatives) thus leaving votes sparse and imbalanced. Consequently, decision-making process yields in undesirable solutions. As a part of solution one can employ probabilistic matrix factorization techniques. As a result, votes are going to be imputed with the most probable values. By having a full voters data matrix more reliable solutions can be obtained.

However, matrix factorization is seldom used for imputation of missing values. More often one uses matrix factorization to investigate crowd characteristics and for validation of the crowd. One such example is presented in paper [5]. Namely, factorization using pBOL method is used for validation of crowdsourced ideas based on expert opinions. The method provides idea-filtering techniques that reduces the number of crowdsourced ideas that will be manually evaluated by experts. This is achieved by creating a predictive model

based on latent features which predict opinion of each expert about the crowdsourced idea. In order to reduce false negatives, the task is transformed from selecting the good ideas to eliminating the poor ones. Compared to pBOL, our framework is used for crowd and expert weighting (instead of filtering), thus allowing automated estimation of importance of crowd votes as well as aggregation of the final solution.

Table 1. Experts vs. crowd – different aspects of collective decision-making [22]

Different aspects		Experts	Crowd
Voting properties [10]:	Selector qualification	Small groups of qualified judges	Many voters with unknown qualifications
	“Selector-selectee” relationship	Experts as an independent body	Voters and candidates can be from the same crowdsourcing community
	Selection process	A systematic voting process with justified evaluation criteria	An uncertain voting process in which participants have their own evaluation criteria
Decision making relies on...		...intuition and reasoning at the same time [11]	...collective intelligence and “wisdom of the crowd” [12]
Application		Complex decision-making problems (e.g., MCDM methods with mutual interdependencies of criteria [13])	Social choice topics with respect to individual preferences as a central topic of AI [14]
Challenges		Possible difficulties in multidisciplinary decision making [15]	Impossibility theorems in Social Choice Theory (e.g., Arrow’s theorem [16], the Condorcet paradox [17])

It is worth to mention SmartCrowd framework proposed by [7] that allows 1) characterization of the participants using their social media posts with summary word vectors, 2) clustering of the participants based on these vectors, and 3) sampling of the participants from these clusters, maximizing multiple diversity measures to form final diverse crowds. They show that SmartCrowd generates diverse crowds and that they outperform random crowds. They estimate the diversity based on external data (tweets). In a sense, this research also tries to estimate diversity of crowds but with respect to both crowd and expert members and without external information.

Expert weighting has also been done in the CI area. One can find such an example in paper [4] where the task was to assign weight to voters for stock pick decisions. This

was done using metaheuristics, namely genetic algorithm. Information about previous judgments and their accuracy as well as additional information (i.e. sentiment analysis from social media) are inserted in genetic algorithm that produces a probability that a crowd voter is an expert. As a result, a framework has a predictive model that can be used for future crowd voters. They showed better average performance than the S&P 500 for two test time periods, 2008 and 2009, in terms overall and risk-adjusted returns. However, this approach assumes existence of historical data (and other additional information) to be available at the predicting model-learning phase and for evaluation of a new crowd voter. In majority of CI examples, one cannot expect to have such an amount of information about crowd voters. Thus it allows weighting and aggregation of crowd and expert votes without collection of additional data. Unsupervised approach seems like an intuitive solution. Unsupervised approach would represent identification of the experts from the crowd voters by using only current votes. We propose one such approach based on similarity matching of experts and crowds. As a result, it is expected to have better decision-making process with greater satisfaction of both crowd and experts.

However, bias in crowd-voting systems can exist. In paper [6], one can find an investigation of the influence of bias in crowd-voting systems with a special focus on ordinal voting. They showed that ordinal rankings often converge to an indistinguishable rating and demonstrated this trend in certain cities for the majority of restaurants to all have a four-star rating. Additionally, they also show that ratings may be severely influenced by the number of users. Finally, they conclude that user bias in voting is not a spam, but rather a preference that can be harnessed to provide more information to users. Based on analyses of global skew and bias they suggest explicit models for better personalization and more informative ratings. Even though research of [6] does not model expert and crowd votes, their research is highly applicable to framework that we are proposing in this paper, because performance of the framework is highly dependent on skew and bias in the data.

3 Framework for Expert-Crowd Voting

Based on opportunities and challenges of crowd voting, as well as potential of benefits of integration of crowd and domain expert knowledge we propose CrEx-Wisdom (Crowd and Expert Wisdom) framework for fusion of experts and crowd “Wisdom” for problems of participatory voting and ranking. The main idea of the framework is to utilize knowledge from a limited number of experts in order to validate and weight crowd votes. Another important aspect that we want to model is the agreement (variance) of both expert and crowd votes as well as their mutual agreement in order to address problems of bias in crowd-voting (described in previous sections). It is important to note that the proposed framework can work in a completely unsupervised manner. This means that expert efforts are reduced to giving opinion on the problem by ranking or grading subset of alternatives, without the need for validation of crowd votes or tracking of crowd voters’ performance history or adding external data. Finally, we try to address the problem of sparsity and aggregation of crowd and expert votes. The data flow of the proposed framework is depicted on Fig. 1.

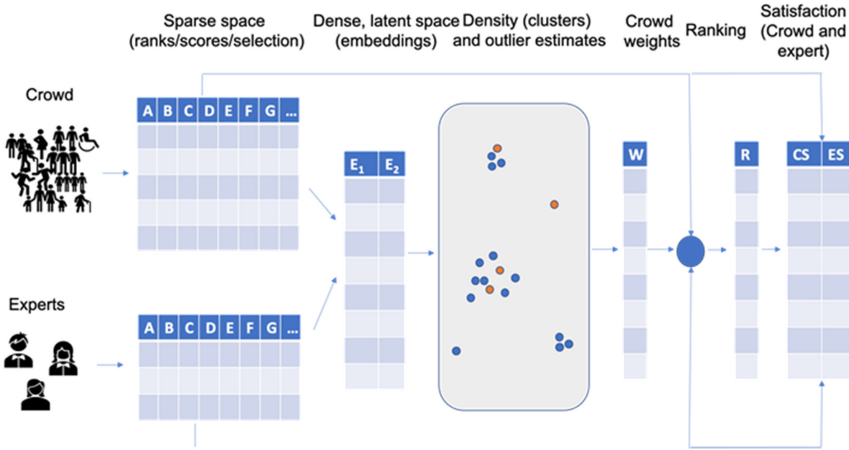


Fig. 1. CrEx-Wisdom framework – data flow

General data and process can be described in following way:

- Experts and crowd are providing votes (ranks, grades, etc.) that are stored in a sparse format.
- Votes of both expert and crowd groups are aggregated in one dataset.
- Latent features (embeddings) are identified based on machine learning (i.e. collaborative filtering) methods.
- Based on latent space of features agreement between experts and crowd (and their mutual agreement) is quantified with machine learning methods such as clustering and outlier detection.
- Based on estimated agreement levels votes of both experts and crowd are weighted on individual level (each voter may have unique weight).
- Votes are aggregated based on traditional methods (e.g. weighted majority) and converted to ranks or grades.
- After aggregation expert satisfaction and crowd satisfaction are measured, and a Pareto front of non-dominated solutions is generated.

CrEx-Wisdom framework provides quite general guidance for fusion of crowd and expert votes in terms of selection of methods and techniques in each step.

In the latent features identification phase, we use the matrix factorization algorithm Alternating Least Squares (ALS) [23] in order to learn latent user and alternative factors. Matrix factorization assumes that each user can be described by k attributes (factors), and each alternative can be described by an analogous set of k attributes (factors). The final prediction (rating) is obtained by multiplication of these two matrices of the voter and alternative factors in order to get a good approximation of missing user ratings. Final model can be represented as (1):

$$\hat{r}_{ui} = x_y^T \cdot y_i = \sum_k x_{uk} y_{ki} \tag{1}$$

where \hat{r}_{ui} represents prediction for the true rating r_{ui} , and $y_i(x_u^T)$ is assumed to be a column (row) vector of user and items called latent vectors of low-dimensional embeddings. Loss function that we used is minimizing the square of the difference between all points in our data (D). Formula of loss function is given in (2):

$$L = \sum_{u,i \in D} (r_{ui} - x_u^T \cdot y_i)^2 + \lambda_x \sum_u \|x_u\|^2 + \lambda_y \sum_u \|y_i\|^2 \quad (2)$$

We also added on two regularization terms in order to prevent overfitting of user and alternative vectors.

ALS algorithm is selected because of cutting edge performance in terms of ranking quality, but also because of its scalability that enables work with big data. Additionally, ALS (and other matrix factorization algorithms) provides convenient representation of both voter and alternative spaces. As such it is important since it allows characterization and application of clustering and/or outlier detection techniques in space of the voters as well as in the space of alternatives.

On the other hand, many different popular techniques may be used e.g. autoencoders, Word2Vec [24], Glove [25], and similar algorithms that showed cutting edge performance in NLP (Natural Language Processing) problems.

Similarly, in this research, we used the well known K-means algorithm [26] (clustering) and Isolation forest [27] (outlier detection) for estimation of voters agreement (density, variance), but we acknowledge that other types of algorithms may be used and possible achieve even better results. However, this investigation is out of the scope of this research since the objective is to show value of integration of crowd and expert votes with machine learning approach.

Considering that the goal of this research is to maximize crowd satisfaction with respect to expert opinion, we used two metrics. The first metric is Satisfaction, which we define as expected value of alternatives number that overlaps with the final decision. This metric does not take into account the ranks of alternatives; we consider that one is satisfied if their favorite alternative is chosen in the first ten ranks. The formula for this metric is given in (3):

$$\text{Overlap}_{wi} = \sum_{j=1}^n (x_{wj} * x_{ij}) \quad (3)$$

$$E(\text{Overlap}) = \sum_{i=0}^{k=10} p(\text{overlap}) * \text{overlap}$$

Where:

n – Number of alternatives (countries, songs);

k – Number of selected (winning) alternatives;

x_{wj} – A boolean value of j-th alternative;

x_{ij} – A boolean value of j-th alternatives for i-th voter.

We considered that the rank difference is more important and in order to capture it, we evaluated our methods using the average points difference from winning combination of alternatives. The formula of this metric is given in (4).

$$\text{avg PD} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n |x_{wj} - x_{ij}| \quad (4)$$

Where:

m – Number of voters;

n – Number of alternatives;

x_{wj} - Winning alternative points at rank j ;

x_{ij} - Alternative points of i -th user at rank j .

4 Experimental Evaluation

In this research, we analyzed the problem of aggregation of crowd and expert votes from the Eurovision song contest. In this contest crowd is represented by televoting participants for each country.

4.1 Data

Votes are aggregated for every country by experts and crowd (televoting), which means, that we had the same number of instances for experts and for crowd. Data used in our experiments are from three years: 2016, 2017, and 2018 for all types of events, which include first-semifinal, second-semifinal, and grand-final. Votes from each country, for both experts and crowd (televoting participants) have a total points range of 58 which is distributed as 1 to 8, 10 and, 12 points. The number of countries that have the right to vote in grand-final is 42 and they can choose from 26 countries that took part in the final contest. In the semifinals, the number of countries that can vote is 21 each and they have the option to choose from 18 available songs. Out of fairness is not allowed for countries to vote for themselves. In a current Eurovision voting setup, the final decision is made by weighting crowd and experts evenly.

4.2 Experimental Setup

We conducted several experiments for different voting methods. In order to compare our methods we used two benchmark methods, one is the current Eurovision weighting method and we created a simple “Single Weighting Crowd” method based on distance from experts.

The Benchmark model that we created is based on weighting voters based on distance from experts. Distance is calculated for each crowd participant to every expert, and then the minimum value is converted to similarity, which represents the weight of a particular voter in the crowd. Every crowd vote is multiplied with its calculated weight, and then the crowd data is summarized together with expert votes in order to get final winning ranking. It is important to note that in this similarity definition latent features (matrix factorization) were not used for representation of voter and alternative space, but rather sparse ranks from original data.

In order to find latent factors (embeddings) of voters and alternative spaces, and consequently define similarities between voters we optimized Alternating Least Squares (ALS) that we trained using Mean Absolute Error. Training is done by splitting data on train and test set. Expert and crowd votes are used together and part of their votes has

been masked and used for measuring error on test set. Several hyper-parameters have been optimized in order to minimize error on test data. Greed search of parameters is shown in the table below (Table 2).

Table 2. Hyperparameter greed search optimization of ALS

Parameter	Grid of values
Number of latent factors	[2, 3, 5, 7, 10, 15, 20]
Regularizations	[0.01, 0.1, 1.0, 10, 100]

After this procedure the best parameters were found, single weight of every crowd is determined as a distance of crowd factor data and expert factor data, which is converted to similarity and used to weigh every crowd vote with corresponding similarity weight.

Further, we tried to identify homogenous groups of experts and describe them with representatives (centroids). These representatives enabled us to simulate situation of much smaller number of experts. Additionally, we used these centroids for measuring similarity with the crowd and assign weights to each crowd participants. Based on exploratory analysis of factor data we saw that there are experts that form homogeneous groups. Hence we used K-means algorithm where we optimize the number of clusters for every data set using Silhouette index as a measure of clusters quality. Here K-means can be replaced with any other cluster algorithm with different measures of quality of detected homogenous groups.

Additionally, we identified outliers in embedded space and conducted the same experimental procedure but with outliers removed from the data.

It can be concluded from description of CrEx-Wisdom framework and experimental setup, crowd votes are weighted based on similarity with expert votes. This means that overall satisfaction of expert voters should increase compared to the current contest voting method (aggregation of expert and crowd votes with equal weights). Therefore, we evaluate the proposed methods in Pareto terms: maximize satisfaction of experts while minimizing “dissatisfaction” of crowd compared to the current voting procedure.

4.3 Results and Discussion

As explained earlier, we used two evaluation metrics one that takes into account only overlapping of selected alternatives with crowd and expert votes and another one that includes rank differences using the number of points given to each rank. Due to space limitation, we will discuss only results of average points difference.

Figure 2 shows percentage of change of voter and expert satisfaction (blue and orange bars, respectively) compared to current voting system:

- for each method (x-axis)
- each event and each competition year (y-axis)

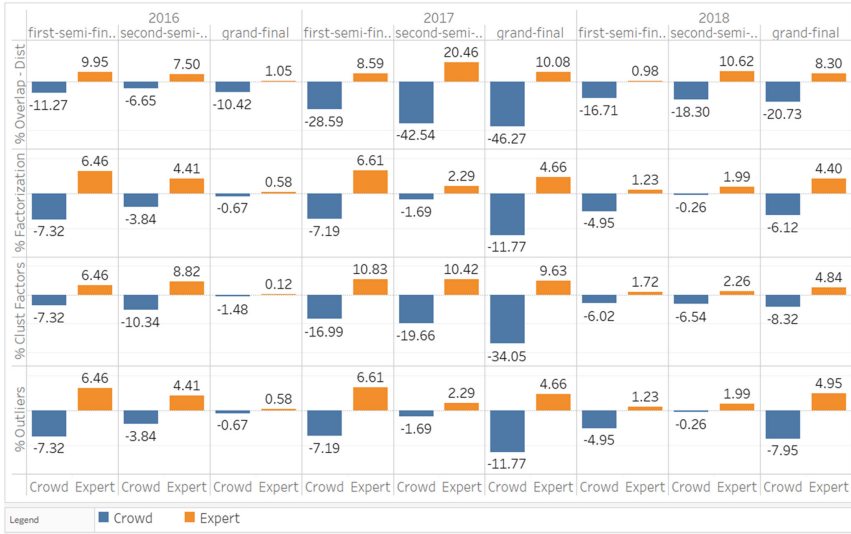


Fig. 2. Relative change in points difference with regard to Eurovision voting (Color figure online)

It can be seen that these changes vary over both years, events and proposed methods.

In order to easier spot differences in performance between methods, the relative change of satisfaction is expressed as ratio of absolute crowd percentage change over absolute expert change (showed in detail in Fig. 2) and presented in Table 3. This ratio practically shows decrease of crowd satisfaction for unit increase of expert satisfaction. Meaning that best results are achieved with minimal values in Table 3.

Table 3. The ratio between crowd and expert change in points

Year	Event type	Overlap	Factorization	Clust – Factors	Outliers
2016	first-semi-final	1.13	1.13	1.13	1.13
2016	grand-final	9.90	1.15	12.64	1.15
2016	second-semi-final	0.89	0.87	1.17	0.87
2017	first-semi-final	3.33	1.09	1.57	1.09
2017	grand-final	4.59	2.53	3.54	2.53
2017	second-semi-final	2.08	0.74	1.89	0.74
2018	first-semi-final	17.00	4.03	3.50	4.03
2018	grand-final	2.50	1.39	1.72	1.61
2018	second-semi-final	1.72	0.13	2.90	0.13

It can be seen from Table 3 that factorization and outlier detection methods are the best performing in most of the cases. However, there are some exceptions. In year 2016

in the first semifinal, we can see that all methods had the same results. We conducted more detailed inspection of embedded data (Fig. 3) that we compressed using T-SNE algorithm in order to visualize points in a two-dimensional space. On these graphs, every point is colored - blue for crowd group, and orange for expert group. It is important to note that the shape represents corresponding cluster labels and that for the convenience of visualization the whole crowd is represented as one cluster (labeled “-1” since only expert data were clustered).

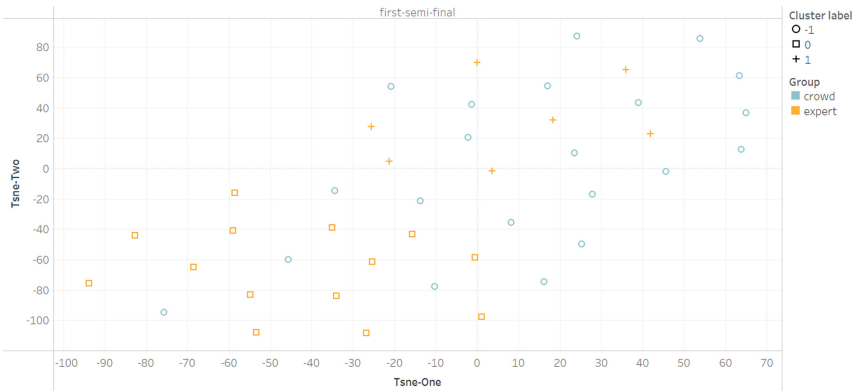


Fig. 3. TSNE 2016 first-semifinal (Color figure online)

Analyzing Fig. 3 we can conclude that the same performance of all voting methods is because of a high dispersion of data. It is clear that there are no homogenous groups neither within expert group or crowd group. Similarly, there are no outliers in this data. On the other hand, in 2016 in grand final, it can be seen that factorization notably outperforms clusters. On Fig. 4 are shown data of this event.

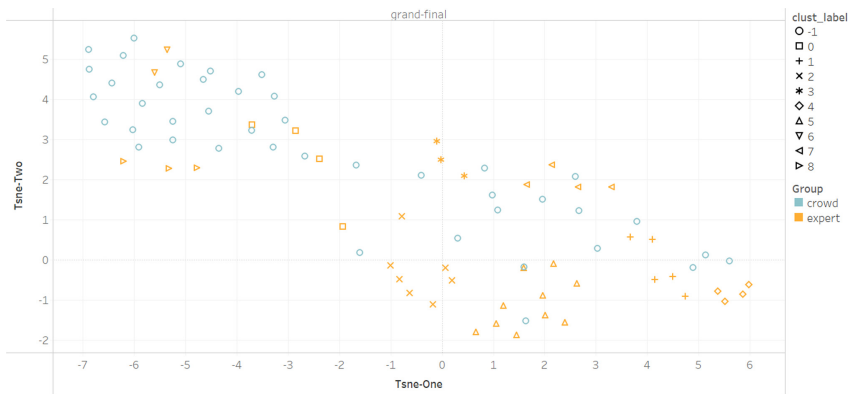


Fig. 4. TSNE grand final in 2016

It can be seen from Fig. 4. That our clustering optimization method found 9 clusters of experts. Such a large number of clusters with respect to number of instances (several clusters have only two or three members) reveals high diversity in expert opinions that is emphasized even more by representing cluster of experts with centroids. We hypothesize that usage of other types of clustering algorithms such as hierarchical clustering could lead to better quality grouping with respect to cluster density. From Fig. 4 it can be seen that most of the experts and a significant number of crowd voters are grouped in the lower right part of the space. This could mean that the final decision (ranking) should be positioned in that part of the space in order to maximize satisfaction of experts and minimize dissatisfaction of crowd.

Additionally, in 2018 in first-semifinal there is a situation where clusters outperform factorization. On Fig. 5 we can see that the cluster algorithm found five quite homogeneous clusters which diminish the variance of expert votes. Based on those groups similarity of crowd is better generalized and thus result from Table 3 is better compared to other methods.

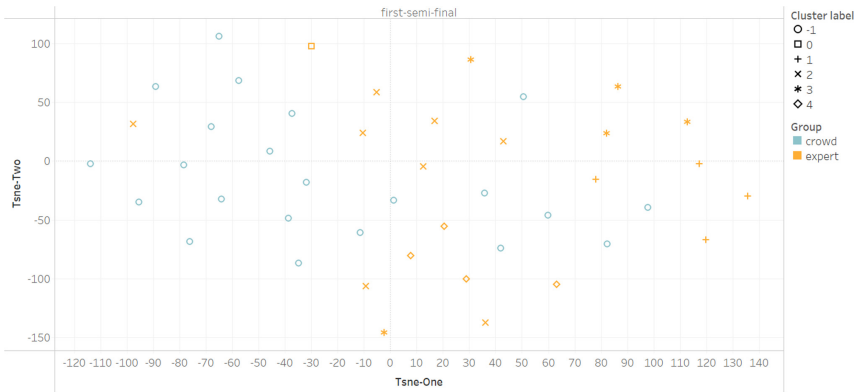


Fig. 5. TSNE 2018 first-semifinal data

In addition, one of the reasons for these results might come from the nature of data used for the experiments. Pop music culture is an area where subjective opinions are highly expected. Moreover, the bias in voting between neighboring countries is present and could be seen from the history of voting. Despite all these unfavorable factors, we showed that in cases where at least part of voters (crowd and/or experts) are homogeneous it is possible to increase crowd/expert satisfaction.

5 Conclusion and Future Research

In this paper, we proposed a framework for integration of expert and crowd votes with the idea of achieving good quality solutions that respect to expert opinion and crowd satisfaction. Results showed that weighting of crowd voters on the individual level, representation of votes in latent space and estimation of consensus level between voters with

clustering and outlier detection procedures can have good impact on finding solutions that compromise between crowd and experts, even if these groups are quite different. In future work, we plan to evaluate more machine learning methods for embedding of votes in latent spaces, clustering and outlier detection. Additionally, we plan to analyze results from this research on theoretical level in terms of voters bias, mutual information between experts and crowd, and densities of crowds and experts. Additionally we plan to validate approach against different voting data (e.g. curriculum creation, best paper awards etc.) where we expect less bias and more consistent voting from experts.

Acknowledgments. This paper is a result of the project ONR - N62909-19-1-2008 supported by the Office for Naval Research, the United States: *Aggregating computational algorithms and human decision-making preferences in multi-agent settings*.

References

1. Traunmueller, M., Fatah, G., Schieck, A.: Introducing the space recommender system: how crowd-sourced voting data can enrich urban exploration in the digital era. In: Proceedings of the 6th International Conference on Communities and Technologies, pp. 149–156 (2013)
2. Jung, H.J., Lease, M.: Inferring missing relevance judgments from crowd workers via probabilistic matrix factorization. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1095–1096 (2012)
3. Hong, H., Du, Q., Wang, G., Fan, W., Xu, D.: Crowd wisdom: the impact of opinion diversity and participant independence on crowd performance. In: Twenty-second Americas Conference on Information Systems (2016)
4. Hill, S., Ready-Campbell, N.: Expert stock picker: the wisdom of (experts in) crowds. *Int. J. Electr. Commer.* **15**(3), 73–102 (2011)
5. Garcia, A.C., Klein, M.: pBOL: an idea filtering method based on negative multi-voting and Pareto aggregation (2017). <http://dx.doi.org/10.2139/ssrn.3175329>
6. Lees, A., Welty, C.: Discovering user bias in ordinal voting systems. In: Companion Proceedings of the 2019 World Wide Web Conference, pp. 1106–1110 (2019)
7. Bhatt, S., Chen, K., Shalin, V.L., Sheth, A.P., Minnery, B.: Who should be the captain this week? Leveraging inferred diversity-enhanced crowd wisdom for a fantasy premier league captain prediction. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, no. 01, pp. 103–113 (2019)
8. Suran, S., Pattanaik, V., Draheim, D.: Frameworks for collective intelligence: a systematic literature review. *ACM Comput. Surv. (CSUR)* **53**(1), 1–36 (2020)
9. Dodevska, Z.A.: Computational social choice and challenges of voting in multi-agent systems. *Tehnika* **74**(5), 724–730 (2019)
10. Chen, L., Xu, P., Liu, D.: The effect of crowd voting on participation in crowdsourcing contests. In: Working paper (2019). 39 pages
11. Bennet, A., Bennet, D.: The decision-making process for complex situations in a complex environment. In: Burstein, F., Holsapple, C.W. (eds.) *Handbook on Decision Support Systems 1*, pp. 3–20. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-48713-5_1
12. Yu, C., Chai, Y., Liu, Y.: Literature review on collective intelligence: a crowd science perspective. *Int. J. Crowd Sci.* **2**(1), 64–73 (2018)
13. Mandic, K., Bobar, V., Delibašić, B.: Modeling interactions among criteria in MCDM methods: a review. In: Delibašić, B., Hernández, J.E., Papathanasiou, J., Dargam, F., Zaraté, P., Ribeiro, R., Liu, S., Linden, I. (eds.) *ICDSST 2015. LNBP*, vol. 216, pp. 98–109. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18533-0_9

14. Rossi, F.: Preferences, Constraints, Uncertainty, and Multi-Agent Scenarios. ISAIM (2008)
15. Jackson, S.E.: The consequences of diversity in multidisciplinary work teams. In: West, M.A. (ed.) *Handbook of Work Group Psychology*, pp. 53–75. Wiley, Chichester (1996)
16. Miller, N.R.: Reflections on Arrow’s theorem and voting rules. *Publ. Choice* **179**(1–2), 113–124 (2019). <https://doi.org/10.1007/s11127-018-0524-6>
17. Herings, P.J.-J., Houba, H.: The Condorcet paradox revisited. *Soc. Choice Welfare* **47**(1), 141–186 (2016). <https://doi.org/10.1007/s00355-016-0950-7>
18. Aitamurto, T., Landemore, H., Galli, J.S.: Unmasking the crowd: participants’ motivation factors, expectations, and profile in a crowdsourced law reform. *Inf. Commun. Soc.* **20**(8), 1239–1260 (2017)
19. Goel, A., Krishnaswamy, A.K., Sakshuwong, S., Aitamurto, T.: Knapsack voting for participatory budgeting. *ACM Trans. Econ. Comput. (TEAC)* **7**(2), 1–27 (2019)
20. Ghezzi, A., Gabelloni, D., Martini, A., Natalicchio, A.: Crowdsourcing: a review and suggestions for future research. *Int. J. Manage. Rev.* **20**(2), 343–363 (2017)
21. Isinkaye, F.O., Folajimi, Y.O., Ojokoh, B.A.: Recommendation systems: P, methods and evaluation. *Egyptian Inform. J.* **16**(3), 261–273 (2015)
22. Dodevska, Z., Kovacevic, A., Vukicevic, M., Delibasic, B.: Two sides of collective decision making - votes from crowd and knowledge from experts. In: *ICDSST 2020, EWG-DSS 6th International Conference on Decision Support System Technology (2020, in press)*
23. Takács, G., Tikk, D.: Alternating least squares for personalized ranking. In: *Proceedings of the Sixth ACM Conference on Recommender Systems (2012)*
24. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013)
25. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, October, pp. 1532–1543 (2014)
26. Lloyd, S.P.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**(2), 129–137 (1982)
27. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: *2008 Eighth IEEE International Conference on Data Mining, 15 December 2008*, pp. 413–422. IEEE (2008)