PhD proposal

Elicitation and explanation for voting rules

Motivation

Voting rules are formal means to aggregate preferences of a group of voters into a collective decision. Social choice theory has produced and analyzed several voting rules, many of which being apparently reasonable: no rule can be considered "the best voting rule", independent of the context. As a result, the choice of the voting rule to be used turns out to be an interesting question. For example, in a (social or commercial) enterprise a body may have to decide how conflicts will be sorted out, or a parliament may decide about a possible revision of the electoral law of the country.

In a given context, there is usually no single best voting rule, but some rules are, upon reflexion, better than others. It is however difficult to see what good properties a voting rule satisfies when only the algorithmic description of its way of aggregating votes is given. Axiomatic analysis is an alternative and well-accepted approach to studying the properties satisfied by voting rules. It consists in formulating "our subjective intuition of fairness (...) as a set of precise desiderata that any acceptable arbitration scheme must fulfil. [The problem is then] reduced to a mathematical investigation of the existence of and characterization of arbitration schemes which satisfy the axioms" (Luce et al., 1957, p. 121).

This research program has revealed extremely successful, with an enormous literature devoted to axiomatic analysis of voting rules (Arrow, Sen, & Suzumura, 2002; Sanver & Selçuk, 2010). However, to the best of our knowledge, scientific literature so far did not investigate rigorous ways of leveraging these findings to effectively help a committee debate about, or choose, a voting rule. It is currently done informally, with committees interviewing experts in social choice and asking them for advices.

There is a need for more rigorous procedures, because knowing which rules satisfy which axioms is not sufficient to solve the problem of choosing a voting rule, but rather transforms it into the problem of knowing which axioms the committee wants to see satisfied. Because all desirable axioms can't be satisfied together, it is not a trivial matter.

The committee can't be simply asked which axioms it likes, because it is often impossible for non-experts (and difficult for experts) to see the consequences of accepting a set of axioms. (For example, a voting rule that satisfies Independence of irrelevant alternatives, Pareto dominance and Universal domain might look appealing to a set of non-experts, if they are not told that this leads to the Dictatorial voting rule.)

Furthermore, the choice of which (set of) axioms should be satisfied should not be left to scientists. But it is unknown how to present axioms and their consequences to a non-expert audience in an understandable way and without unduly influencing it.

Finally, the choice of a voting rule can ultimately rest, in some contexts, on subjective preferences of a committee. In this case, it is important to develop elicitation techniques that can capture those preferences in some formal language. Elicitation techniques are currently seldom applied to preference about abstract objects such as voting rules (see below).

Related research

There is an interesting literature in social choice theory on the analysis of a group of voters

deciding on which voting rule to use (Koray, 2000; Barbera et al., 2004), where voters have a basic preference over alternatives and they compare any two voting rules according to the outcome they induce. This approach overlooks any value judgment of voters regarding the axioms that voting rules could satisfy, but still, might provide fruitful interactions with our focus. There is also a recently initiated research by Jean Lainé and Remzi Sanver who address the same question, now with voters directly having preferences over axioms. Clearly, their model is much closer to the problem we aim to handle.

There is currently a big attraction towards explanation of recommendations, both in machine learning (Lipton, 2017) and in multiple criteria decision aiding. This can be seen as driven by an increasing desire of society for justification of science-driven (often opaque) recommendations. Some of these techniques, of special interest to us, rest on models that represent formally the preference of the user receiving the recommendation (Belahcene et al., 2017).

General elicitation procedures using active learning are much developed currently (Fürnkranz & Hüllermeier, 2010), and there is an interest in the literature for obtaining explainable and simple preference models. Techniques for querying for preferences about voting rules (Cailloux et al., 2014) can be combined to these approaches (we started a research work recently about this, involving Olivier Cailloux, Paolo Viappiani and Stefano Moretti).

Work aiming at deducing automatically properties of voting rules exist already (Geist et al., 2011) and show the interest of combining axiomatic analysis and computer science approaches. We started a research work (involving Michael Kirsten and Olivier Cailloux) which uses automated theorem proving techniques to detect automatically, given a voting rule and a set of axiom specified in some formal language, which axioms the voting rule satisfies.

The problem of debating about voting rules can also be seen as one of selecting the best arguments, which is studied in the formal argumentation theory community, and has been recently applied to the voting rule context (Cailloux et al., 2016). Another current work aims at indicating sufficient properties that guarantee that “enough” arguments have been considered to ensure a valid recommendation (Cailloux et al., 2017).

Expected results and impact

The phd work will aim at publishing work in AI conferences and journals and in the social choice field. It may contribute to, or benefit from, the research works presented above. It will happen in a context favorable to the development of “white-box”, explainable models, as detailed above; and may contribute to make scientific-driven recommendations more acceptable to the general public.

Besides theoretical work, it will aim at producing practical methods that could be integrated into the Whale platform, which would permit to give a practical twist to the research.

As an example, automatic deduction might be used to generate profiles that illustrate various (paradoxical or intuitive) results when applying known voting rules to concrete situations. This shows the potential usefulness of those approaches beyond the original motivation of this project.

Expected assets

The subject involves social choice, preference modeling, formal argumentation theory. The candidate must know well some of these fields, and be interested in the other ones.

Contacts

Supervisor: Remzi Sanver. Co-supervisor: Olivier Cailloux.

References

Arrow, K. J., Sen, A., & Suzumura, K. (Eds.). (2002). *Handbook of social choice and welfare*. Elsevier.

Barbera, S., & Jackson, M. O. (2004). Choosing How to Choose: Self-Stable Majority Rules and Constitutions. *The Quarterly Journal of Economics*, *119*(3), 1011–1048. https://doi.org/10.1162/0033553041502207

Belahcene, K., Labreuche, C., Maudet, N., Mousseau, V., & Ouerdane, W. (2017). Explaining robust additive utility models by sequences of preference swaps. *Theory and Decision*, *82*(2), 151–183. https://doi.org/10.1007/s11238-016-9560-1

Cailloux, O., & Endriss, U. (2014). Eliciting a Suitable Voting Rule via Examples. In T. Schaub, G. Friedrich, & B. O'Sullivan (Eds.), *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI 2014)* (Vol. 263, pp. 183–188). IOS Press. https://doi.org/10.3233/978-1-61499-419-0-183

Cailloux, O., & Endriss, U. (2016). Arguing about Voting Rules. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)*. IFAAMAS. Retrieved from http://www.ifaamas.org/Proceedings/aamas2016/

Cailloux, O., & Meinard, Y. (2017). A formal framework for deliberated judgment. Retrieved from https://arxiv.org/abs/1801.05644. Submitted to Theory and Decisions

Fürnkranz, J., & Hüllermeier, E. (Eds.). (2010). *Preference Learning*. Springer. Retrieved from http://doi.org/10.1007/978-3-642-14125-6

Geist, C., & Endriss, U. (2011). Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *Journal of Artificial Intelligence Research*, *40*, 143–174. https://doi.org/10.1613/jair.3126. IJCAI-JAIR Best Paper Prize 2016

Koray, S. (2000). Self-Selective Social Choice Functions Verify Arrow and Gibbard-Satterthwaite Theorems. *Econometrica*, *68*(4), 981–996. https://doi.org/10.1111/1468-0262.00143

Lipton, Z. C. (2017). The Mythos of Model Interpretability. Retrieved from https://arxiv.org/abs/1606.03490. Presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY

Luce, R. D., & Raiffa, H. (1957). *Games and Decisions*. J. Wiley, New York.

Sanver, M. R., & Selçuk, Ö. (2010). A characterization of the Copeland solution. *Economics Letters*, *107*(3), 354–355. https://doi.org/10.1016/j.econlet.2010.03.002