Summary of the work entitled ***'Calculating the interaction index: a polynomial approach based on sampling'***

Inmaculada GUTIÉRREZ

joint work with Javier CASTRO, Daniel GÓMEZ, and Rosa ESPÍNOLA

1. Context

In the second half of the 20^{th} century, the measure theory [1] was increasingly popular. It has had a great deal of importance in the develop of several areas, as the fuzzy measures [2], the necessity and possibility measures [3, 4], the capacity measures [5], etc.

We focus on fuzzy measures. These monotonic set functions define a type of non-additive measures encompassing many measures such as belief, possibility, necessity or plausibility measures [6]. Fuzzy measures have been widely analyzed [7, 8] and applied in many fields. Unfortunately, the practical implementation of fuzzy measures is complex (2^n coefficients are needed to define a fuzzy measure over n individuals), and the related semantic is difficult to understand [9]. Many researchers have focused their attention on the characterization, interpretation and representation of fuzzy measures, which may be understood as some specific cooperative games for which monotony holds.

In an attempt to understand fuzzy measures, it turns essential the analysis of the interactions between the individuals. As extension of the Shapley value [10], it was defined the interaction index [11], axiomatically characterized to measure the interaction phenomena between individuals regarding a fuzzy measure, players in cooperative game theory, or criteria in multicriteria decision analysis [12]. As a generalization of this index, it was defined the representation index to cope with the relations of any set. In this work we deeply analyze and understand these indices, providing an alternative and simpler characterization of them as well as a novel methodology to estimate their real value (which is a $NP-$problem in general scenarios).

2. State of Art

Many researchers have focused their attention on the characterization, interpretation and representation of fuzzy measures and the interaction of the individuals Probably, the first attempt to the analysis of interactions is due to Owen [13], followed by the works of Roubens [14] and Marichalar and Roubens [15]. The notion of interaction index, that may be understood as an extension of value [16–19] is essential to measure and interpret the relations between the individuals regarding a fuzzy measure, and it may refer to both redundancy and complementary effects between players or coalitions.

It is worth highlighting the importance index Shapley value [10], an essential tool in game theory which has been adapted to the field of fuzzy measures, in which it could be understood as an indication of the importance of each singleton. This index has been deeply studied from a theoretical point of view [20, 21]. Nevertheless the calculation of its real value for general fuzzy measures is a NP-problem. Until not that long ago, there were only a few works to approximate this index in some specific problems, for example, when the population has some particular properties [22, 23]. In 2009, Castro et al. proposed a method based on sampling to approximate the Shapley value [24]. It was generalized in [25] by a stratified random sampling with optimum allocation process.

It is also interesting to analyze the interactions among several elements [26]. In 1993, Murofushi and Soneda defined the interactions index between two individuals [11], which quantifies its average contribution considering all the subsets it is part of. Then, in 1997, Grabisch proposed an extension of this index to deal with sets involving more than two elements [27]. As with the Shapley value, the approximation of the real value of these indices has not been studied in depth for general scenarios, due to its great complexity.

We propose an alternative characterization based on orders of this interaction index, for both the simple case related to pairs of elements, and the corresponding extension to deal with any set. This new formulation of the interaction index (and of the representation index), provides a different and intuitive understanding of this it. Then, following the idea of sampling in [24, 25], we define two methods to approach the value of the interaction index. Both are based on sampling, where the second one is a refinement of the very first, but also using stratified random sampling with optimum allocation.

3. Main Contributions

• Proposal of an alternative characterization of the interaction index based on orders. Murofushi and Soneda [11] introduced an interaction index $I_{ij}(\mu)$ based on several concepts related to multi-attribute utility theory to model the interactions between pairs of elements i, j regarding the fuzzy measure μ, i.e. to estimate the degree to which a pair of elements interact. $I_{ij}(\mu)$ can be seen as an average of the added value obtained by considering i and j in the same coalition, depending on μ. In fact, considering a singleton, the Shapley value coincides with the interaction index. We work in an alternative characterization of the interaction index based on orders and permutations, which provides a simpler view of it.

On the other hand, in [27], by analogy with the first- and second-order classes, Grabisch proposed a generalization of I_{ij}, $I_T(\mu)$ to represent interactions among the elements of any set T regarding the fuzzy measure μ. This index, named representation index, is useful to quantify the average contribution of a coalition when considering all the subsets it is part of. Once again, for $T = \{i\}$, recovers $Sh_i(\mu)$, and I_{ij} for $T = \{i, j\}$. As with I_{ij}, we propose an alternative characterization of I_T based on orders.

These characterizations provide us a new view of the interaction indices (for pairs of individuals or bigger sets) as an average of the added value obtained by considering them in the same coalition, depending on the involved fuzzy measure.

• Development of a new methodology to approach the real value of the interaction index (and the representation index). The complexity of Shapley value calculation is well-known: it is a $NP-$problem for general fuzzy measures (there cases several specific cases for which the computation process is simpler). Although this index has been deeply analyzed from a theoretical perspective, from the best of our knowledge, until not that long ago, there were only a few works that propose an approach to approximate its real value. Furthermore, these references are focused on some specific problems until the proposals in [24, 25].

This lack of practical analysis of the Shapley value also occurs with the interaction index. Its calculation can be done in polynomial time only for a few simple fuzzy measure, and there is not much research about feasible estimations of it. Inspired by [24, 25], we propose a new polynomial-time methodology to approach the interaction index, developed on the basis of its alternative characterization based on orders. We define two polynomial-time algorithms to estimate the real value of the interaction index. The first one, named *ApproInteraction*, is based on simple random sampling. In the full work we detail the step-by-step process and provide its pseudo-code for easy adaptation to any programming language. The obtained estimation has some desirable properties, specifically, it is unbiased and consistent in probability.

There are some situations in which the *ApproInteraction* algorithm may not be accurately enough, because of the large variance caused by simple random sampling, which can be reduced with strata. We propose a refinement of the *ApproInteraction* algorithm with the use of stratified random sampling named *StratifiedApproInteraction*, which improves the performance of the original method for cases with high variance. The key is to divide the whole population into subpopulations, each of which is internally homogeneous and as heterogeneous as possible with respect the others groups. As with the simple random sampling method, in the full work we show a detailed explanation of the *StratifiedApproInteraction* algorithm, as well as its pseudo-code. Again, the estimated value has desirable statistical properties, as being unbiased and consistent in probability.

Because of the lack of methods to estimate the interaction index, we can not compare our algorithms with others. Then, we calculate the real value of I for several simple fuzzy measures for which this calculation can be done in polynomial-time (considering the alternative characterization based on orders), and we compare the real value with the estimations. Both methods are so accurate, so the obtained results are exact and reliable, so we can confirm the goodness of our methodology.

To wrap up this point, note that although we thoroughly explain and test the methods with respect to the interaction index, the stratified procedure can be generalized in much the same way to estimate other values, specifically for the representation index I_T. In fact, the process is summarized by the corresponding pseudo-code in the full text, named *Stratified ApproInteraction* T algorithm.

Although we have focused on fuzzy measures, all the contributions may be adapted to a more general scenario related to cooperative game. We can state it because fuzzy measures can be understood as a subset of cooperative games with monotony.

All these contributions have culminated in the elaboration of a research paper, submitted to the high impact journal *Fuzzy Sets and Systems* to be published. At the time of presenting this candidature, the paper is in the reviewers phase. We hope to receive good news about its publication soon. Furthermore, all the results are included in the doctoral thesis *'Detección de comunidades en redes mediante el uso de medidas borrosas'*, and considered to develop some results in several papers, as [28, 29]

References

[1] P. Halmos, Measure Theory, Springer, 1950. doi:https://doi.org/10.1007/978-1-4684-9440-2.

[2] M. Sugeno, Fuzzy measures and fuzzy integrals: A survey, Fuzzy Automata and Decision Processes 78 (1977) 89–102.

[3] L. A. Zadeh, Similarity relations and fuzzy orderings, Information Sciences 3 (1971) 177–200. doi:10.1016/S0020-0255(71)80005-1.

[4] L. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28. doi:10.1016/S0165-0114(99)80004-9.

[5] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1954) 131–295. doi:10.5802/aif.53.

[6] M. Grabisch, T. Murofushi, M. Sugeno, Fuzzy measure of fuzzy events defined by fuzzy integrals, Fuzzy Sets and Systems 50 (3) (1992) 293–313. doi:10.1016/0165-0114(92)90227-U.

[7] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, Patter Recognition Letters 17 (6) (1996) 567–575. doi:10.1016/0167-8655(96)00020-7Getrightsandcontent.

[8] G. Beliakov, J. Wu, Learning k-maxitive fuzzy measures from data by mixed integer programming, Fuzzy Sets and Systems 412 (2021) 41–52. doi:10.1016/j.fss.2020.04.015.

[9] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Amsterdam, The Netherlands, 1980.

[10] L. Shapley, A value for n-person games, Contribute. Theory Games 2 (28) (1953) 307–317. doi:10.1515/9781400881970-018.

[11] T. Murofushi, S. Soneda, Techniques for reading fuzzy measures (III): interaction index, in: Proc. of 9^{th} Fuzzy System Symposium, 9^{th} Fuzzy System Symposium, Sapporo, Japan, 1993, pp. 693–696.

[12] M. Grabisch, M. Roubens, An axiomatic approach to the concept of interaction among players in cooperative games, International Journal of Game Theory 28 (4) (1999) 547–565. doi:10.1007/s001820050125.

[13] G. Owen, Multilinear extensions of games, Management Science Series B-Application 18 (5) (1972) 64–79. doi:10.1287/mnsc.18.5.64.

[14] M. Roubens, Interaction between criteria and definition of weights in MCDA problems, 44th Meeting of the European Working Group *Multiple Criteria Decision Aiding*, 1996.

[15] J. Marichal, M. Roubens, The chaining interaction index among players in cooperative games, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 69–85. doi:10.1007/978-94-017-0647-6_5.

[16] P. Dubey, A. Neyman, R. Weber, Value theory without efficiency, Mathematics of Operations Research 6 (1981) 122–128. doi:10.1287/moor.6.1.122.

[17] J. Banzhaf, Weighted voting doesn't work: A mathematical analysis, Rutgers Law Review 19 (1965) 317–343.

[18] R. Weber, Probabilistic values for games, in: A. Roth (Ed.), The Shapley Value: Essays in Honor of Lloyd S. Shapley, Cambridge University Press, 1978, pp. 101–119.

[19] J. Harsanyi, A simplified bargaining model for the n-person cooperative game, International Economic Review 4 (2) (1963) 194–220. doi:10.1007/978-94-017-2527-9_3.

[20] M. Ballasote, C. Hernández-Mancera, A. Jiménez-Losada, A new Shapley value for games with fuzzy coalitions, Fuzzy Sets and Systems 383 (2020) 51–67. doi:10.1016/j.fss.2018.12.018.

[21] J. M. Gallardo, A. Jiménez-Losada, A characterization of the Shapley value for cooperative games with fuzzy characteristic function, Fuzzy Sets and Systems 398 (2020) 98–111. doi:10.1016/j.fss.2019.10.001.

[22] W. Cochran, Sampling Techniques, Wiley, 1977. doi:10.1002/bimj.19650070312.

[23] H. Lohr, Sampling: Design and Analysis, Duxbury, 1999. doi:10.1201/9780429296284.

[24] J. Castro, D. Gómez, J. Tejada, Polynomial calculation of the Shapley value based on sampling, Computers & Operations Research 36 (5) (2009) 1726–1730. doi:10.1016/j.cor.2008.04.004.

[25] J. Castro, D. Gómez, E. Molina, J. Tejada, Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation, Computers & Operations Research 82 (2017) 108–188. doi:10.1016/j.cor.2017.01.019.

[26] I. Kojadinovic, A weight-based approach to the measurement of the interaction among criteria in the framework of aggregation by the bipolar Choquet integral, European Journal of Operational Research 179 (2) (2007) 498–517. doi:10.1016/j.ejor.2006.03.044.

[27] M. Grabisch, k-order additive discrete fuzzy measures and their representation, Fuzzy Sets and Systems 92 (2) (1997) 167–189. doi:10.1016/S0165-0114(97)00168-1.

[28] I. Gutiérrez, D. Gómez, J. Castro, R. Espínola, Fuzzy measures: a solution to deal with community detection problems for networks with additional information, JIFS 39 (5) (2020) 6217–6230. doi:10.3233/JIFS-189091.

[29] I. Gutiérrez, D. Gómez, J. Castro, R. Espínola, Multiple bipolar fuzzy measures: an application to community detection problems for networks with additional information, IJCIS 13 (1) (2020) 1636–1649. doi:10.2991/ijcis.d.201012.001.