Expert Systems with Applications 39 (2012) 4250–4257

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A decision support system for animated film selection based on a multi-criteria aggregation of referees' ordinal preferences

S. Jullien-Ramasso, G. Mauris *, L. Valet, Ph. Bolon

LISTIC, Polytech Annecy-Chambéry, Université de Savoie, BP 80439, 74944 Annecy le vieux, France

ARTICLE INFO

Keywords:
Animated film selection
Ordinal scale
Expert preference representation
Multi-criteria aggregation

ABSTRACT

This paper presents a decision support system devoted to the selection of films for the International Animated Film Festival organized at Annecy, France, every year. It deals with the representation and aggregation of referees' preferences along predefined criteria in addition to their overall selection point of view. The practical requirements associated with this application (often encountered in social or cultural areas as well) are: a common ordinal scale for the criteria scores, a procedure to deal with inconsistencies between criteria and overall scores, explanation tools of each referee's preference model in order to facilitate the deliberation process and also to argument the selection decision. The processing of referees' preferences is achieved thanks to a recent method which consists in finding a generalized mean aggregation operator representing the preferences of a referee, in a finite ordinal scale context. The method allows to deal with consistency conditions on referees' behaviour in order to highlight the criteria or pair of criteria which are the most influential for each of the referees. All the functionalities have been implemented in an interactive decision software that facilitates a shared selection decision. Results issued from the 2007 selection are presented and analysed from the preference representation and processing point of view.

1. Introduction

The real-sized application considered in this paper is the selection of films for the International Animated Film Festival organized every year at Annecy, France. The board decision making appointed by the festival organizing committee is based on the three referees' preferences along predefined criteria (Scenario, Aesthetics, Animation Technics and Soundtrack) in addition to their overall selection point of view (yes, perhaps, no) on each of the 600 films they watch in one week (100 films per day). The practical requirements associated with this application, but encountered also in other applications such as student evaluation or scientific paper selection, are:

- a common ordinal scale for the criteria scores (e.g. bad, medium, good), avoiding any numerical conversion in order to faithfully represent the revealed preferences of the referees,
- a procedure to deal with inconsistencies between criteria and overall scores due to the limited accuracy and reliability of referees or generated by the very demanding conditions of the evaluation task,
- explanation functionalities of each referee preference model, as the three models are generally conflicting, in order to facilitate the deliberation process and also to argue the selection decision with the stakeholders.

Note that in this film evaluation problem, there are in fact two aggregation levels: aggregation of criteria scores for a given referee to get an overall score, and aggregation of referees' evaluations to get a selection decision. According to the committee requirement, only the former aggregation level has been formalized. The latter does not receive a formal treatment: it is solved in an implicit way by the use of a decision support interface making it possible for the committee to have an idea of the distribution of grades over referee. It provides also each referee with some feedbacks on his internal judgment consistency (Hahn, 2010), and with a compact representation of the value system of each other under the form of qualitative influence indexes attached to each subset of criteria. These decision support functionalities aim at facilitating discussion among referees by providing information for preference analysis in order to lead to a shared decision with some legitimizations (Amgoud & Prade, 2009; Belton & Pictet, 1997). The choice of not formalizing this second aggregation level is motivated by the organizing committee requirement to respect the referees' opinions by letting them the possibility to have their own different value system, but also by the difficulty of the associated information

* Corresponding author.
E-mail address: mauris@univ-savoie.fr (G. Mauris).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.09.109

processing. Indeed, the aggregation of overall scores given by all the referees require that these scores are seen as commensurate, this requires substantial efforts in the construction procedure of these scores.

Concerning the first aggregation level, it could be considered by conventional weighted mean of numerical scores. The problem is that these numerical scores are difficult to elicit directly due to the human inability to have a precise perception leading to the identification of definite scores. Moreover, once they are apparently clearly intelligible to each stakeholder, numerical scores do not always have a sound measurement theory based meaning (Roberts, 1979) in terms of significance covered by the numerical values; e.g. the differences or the ratios are considered to have a meaning though the referees' preferences are purely ordinal. However, there are a few methods to obtain numerical scores from qualitative preferences in a sound way. The Macbeth method (Bana e Costa & Vansnick, 1997) converts preferences into a numerical scale in a way which is coherent with the measurement theory but it requires an intensity of preference from the referees.

The Tomaso method (Marichal, Meyer, & Roubens, 2005) provides numerical information. However, the final score is on a scale which has no relation with the original score scale, which makes it difficult to draw explanations for the stakeholders. One adapted way to obtain a clear and common elicitation of preferences regarding criteria and overall scores is to consider verbal terms on an ordinal scale. Thus, without fictitious numerical conversions, the evaluations preserve the original concrete point of view of each referee who is completely free to behave as he or she feels. In fact, this raises problems for the aggregation of these ordinal evaluations. Indeed, they are a lot of methods in multi-criteria decision making able to deal with ordinal scores such as Electre-like methods (Figueira, Greco, & Ehrgott, 2005), but they produce a rank of the alternatives and not a score as final result. Rule-based approaches can also be considered (Greco, Matarazzo, & Slowinski, 2005; Zhang & Chu, 2009). They rather map alternatives to some predefined ordered categories, which could be considered as scores, but they have to be predefined. Moreover, only a few studies have been done in the domain of aggregation functions on finite ordinal scales and associated explanation capabilities have not been developed. Nevertheless, a recent method proposed by Grabisch allows to deal with the above practical preference processing challenges (Grabisch, 2006). His proposition consists in finding all the mean aggregation operators representing the referee preferences in a finite ordinal scale score context. By definition, a mean operator (in the wide sense) satisfies: internality (i. e. the overall evaluation should not be beyond the range of scores), and non decreasingness (that ensures that an improvement of one score cannot decrease the overall score). Thus this method allows to deal with consistency conditions on referees' score profiles. Moreover, it also provides some responses to the questions raised by a stakeholder in order to obtain a better mutual understanding and thus legitimization of the final decision which is an important issue for the organizing committee.

In summary, the paper presents a decision-making support system for animated film selection by using Grabisch's method for the representation and aggregation of referees' preferences. Section 2 describes the film evaluation procedure, i.e. the considered criteria, the finite ordinal scale scoring, and the determination of a mean aggregation operator. In Section 3, explanation functionalities about the referees' behaviour are proposed in order to highlight the criteria or pair of criteria which are the most influential for each of the referees. Section 4 describes the developed interactive decision software which facilitates a shared selection procedure. Experimental results issued from the 2007 selection are presented in Section 5 and analysed from the preference representation and processing point of view.

2. Film evaluation procedure

In this section, the previous selection procedure used by the organization committee of the Annecy International Animated Film Festival is discussed. We then present the proposed procedure based on predefined criteria with associated scoring, and on Grabisch's method for the determination of a mean aggregation operator identifying some consistent referee types of behaviours.

2.1. Background on the film selection issues

The International Animated Film Festival has been taking place in Annecy in the French Alps every year for the last fifty years. It rewards the best animated films of different formats: short films, full-length films, school films, web films. Before the festival, a selection is made to decide which films will be shown. A cosmopolitan jury composed of three referees with different nationalities and different professional backgrounds makes the selection. The three referees have to watch approximately 600 films of the short film format in order to compare their quality. They then select about 100 films for that format considered in this paper. The number of films to be viewed and the time allotted for that viewing make it a demanding activity for the referees who have to watch 100 films per day during a week. Naturally, the selection process is supervised by the organization committee. Previous selections Jullien, Mauris, Valet, and Bolon (2006), i.e. before 2007, were based on the overall evaluations of the three referees expressed on a four-level scale: *Yes sure*, *Yes*, which meant that the film had been respectively strongly selected and selected, *Perhaps* and *No*. Then, after one screening day, the referees compared their overall evaluations and tried to obtain a consensus on the list of selected films. However, this step was long, due to the fact that different kinds of criteria which had not been explicitly evaluated came in the discussion, additionally the importance given to these criteria by each referee is not the same due to the various referee backgrounds (technical, artistic, media). Therefore, the selection decision often ended by a vote which did not wipe away the conflicts between the referees and, the decision was also difficult to explain to the film submitters and to the media. Therefore, to reduce deliberation time, to facilitate a shared decision, and to be able to legitimate the selection decision with respect to the film submitters and the media, the organization committee of the animated film festival has decided to improve the selection procedure. In this view, the organization committee decided to explicitly incorporate multiple criteria preferences because different criteria can provide a comprehensive and measurable representation of the referees' points of view. Four main criteria were selected in addition to the overall evaluation: *Scenario*, *Aesthetics*, *Animation Technics* and *Soundtrack*. These criteria are well known to the referees and cover the main characteristics of animated films. Thus they allow to develop a shared view of a film description. The overall evaluation is kept for each referee in order he/she does not have the feeling of being excluded from the final decision. Indeed, directly merging all the referees' criteria evaluations to deduce the selection has not been considered. This merging strategy seems like the referees' overall opinion does not really matter and thus incline the referees to over evaluate the criteria as they are not aware of the importance of each criterion. On the other hand, the proposed strategy allows each referee to use its own priorities between criteria, and leaves the room to identify the preference behaviour of each referee useful to guide the final decision debate. Indeed, it is well admitted that, except for a few films, the referees' behaviour is of a generalized mean type, i.e. if a film is better than another on one criterion it cannot be lesser on the overall evaluation which is moreover a trade-off between well rated and poorly rated crite-