

Ordinal classification/regression for analyzing the influence of superstars on spectators in cinema marketing

Elena Montañés^{a,*}, Ana Suárez-Vázquez^b, José Ramón Quevedo^a

^a*Artificial Intelligence Center. University of Oviedo in Gijón. Campus de Viesques 33204 Gijón, Spain*

^b*Faculty of Business and Economics. University of Oviedo in Oviedo. Avda. Del Cristo s/n 33071 Oviedo, Spain*

Abstract

This paper studies the influence of superstars on spectators in cinema marketing. Casting superstars is a common risk-mitigation strategy in the cinema industry. Anecdotal evidence suggests that the presence of superstars is not always a guarantee of success and hence, a deeper study is required to analyze the potential audience of a movie. In this sense, knowledge, attitudes and emotions of spectators towards stars are analyzed as potential factors of influencing the intention of seeing a movie with stars in its cast. This analysis is performed through machine learning techniques. In particular, the problem is stated as an ordinal classification/regression task rather than a traditional classification or regression task, since the intention of watching a movie is measured in a graded scale, hence, its values exhibit an order. Several methods are discussed for this purpose, but Support Vector Ordinal Regression shows its superiority over other ordinal classification/regression techniques. Moreover, exhaustive experiments carried out confirm that the formulation of the problem as an ordinal classification/regression is a success, since powerful traditional classifiers and regressors show worse performance. The study also confirms that talent and popularity expressed by means of knowledge, attitude and emotions satisfactorily explain superstar persuasion. Finally, the impact of these three components is also checked.

Keywords: Ordinal classification, ordinal regression, machine learning, cinema marketing

1. Introduction

In recent years, the movie industry has become a very important issue in economic literature (Marshall et al., 2013). This is a consequence of its relevance both from a quantitative and qualitative point of view. According to the MPAA (Motion Picture Association of America) global box office for all movies released around the world reached 34.7 billion in 2012, up 6% up on the 2011 total. At the same time, cinema screens increased by 5% worldwide in 2012. The movie industry has been characterized as a giant ongoing laboratory in which both consumers and producers conduct experiments (Sedgwick and Pokorny, 2012). While the costs of these experiments for consumers are the ticket price and an opportunity cost, the costs for producers are highly significant (Eliashberg et al., 2006). Characteristics such as the astronomic investments (Ravid, 1999), short life cycles (Jedidi et al., 1998) and an inherent difficulty in assessing the quality of motion pictures (Ginsburgh and Weyers, 1999) confer a high risk to the movie industry.

The movie industry uses different hallmarks of quality in an attempt to control the level of uncertainty. A cast of superstars is one of the most important resources in this sense (Albert, 1999). Anecdotal evidence, however, suggests that a cast of superstars is not a guarantee of box office success. Besides, research on the influence of superstars on box office revenue has yielded somewhat mixed results (Nelson and Glotfelty, 2012). Apart from this lack of consistency in previous results, the majority of the studies analyze the relationship between the presence of superstars and box office performance. However, there is insufficient research about how it could be possible to improve this relationship. This study addresses this gap by analyzing factors that can increase superstars influence on spectators. The development of a behavioral model capable of validating certain factors in superstar influence can be a helpful tool to diminish the risk of a movie. Furthermore, it can be very useful to guide the appropriate actions to enhance the value of a cinema star.

The main objective of this paper involves the development of a machine learning model capable of quantifying superstar influence on spectators. Several factors related

*Corresponding author: E. Montañés (elena@aic.uniovi.es). Phone/Fax: +34 985182645/985182125
Email addresses: elena@aic.uniovi.es (Elena Montañés), anasv@uniovi.es (Ana Suárez-Vázquez), quevedo@aic.uniovi.es (José Ramón Quevedo)

to talent and popularity which are a way of considering movigoers persuasion are studied for this purpose. It has not been until quite recently that machine learning methods have been applied to analyze marketing problems (Abernethy et al., 2008). The applications are even scarcer in the cinema sector and they are only referred to movie recommender systems (Cheung et al., 2003). To the best of our knowledge, no previous application of machine learning methods exists in the field of superstars influence. Special care will be taken in the study of the nature of the data in order to identify the machine learning problem that leads to choose both an adequate formulation and adequate techniques. In any case, once the formulation is adequately stated, an exhaustive set of experiments with different non-ordinal and ordinal techniques will be carried out in order to find out which ones obtain promising models. Besides, additional experiments are performed in order to analyze the influence the different factors related to talent and popularity on their own have on moviegoers' persuasion.

The rest of the paper is organized in different sections. Section 2 describes the features considered to quantify the moviegoers persuasion and how the data were collected. Section 3 details the machine learning problem formulation and proposes and discusses the methods to solve it. Section 4 shows the model extraction process, the experimental settings and results and discusses them. Finally, section 5 outlines the conclusions and some ideas for future research.

2. Materials: Features for quantifying moviegoers persuasion

This section describes the features for quantifying the moviegoers' persuasion in terms of talent and popularity, and hence, it also describes the way the data were collected.

There are many reasons to justify the presence of stars in the cast of a movie (Suárez-Vázquez, 2011): (a) their acting skills increase the quality of the movie; (b) they are an informational signal and (c) they can reduce the effect on the market of unfavorable criticism. The idea behind these explanations is that the presence of a superstar increases the intention of watching a certain movie. In social psychology this type of interpersonal influence is known as persuasion (Cacioppo et al., 1997). Persuasion can be the result of central aspects related with the persuasive object or it can be the product of peripheral cues (Lord et al., 1995). In the case of superstars, the central aspects are related with

their talent (Rosen, 1981) while the peripheral aspects are a result of their popularity (Adler, 2006). Previous studies have shown that talent and popularity are imperfectly correlated (Filimon et al., 2011). This duality in stardom facets does not show a lack of rationality in spectators' behavior. On the contrary, it indicates that spectators' behavior is a complex phenomenon and many variables should be taken into account to approximate it.

The influence of different persuasive evidences has been extensively studied in the field of advertising (Hansen et al., 1997). However, there are no applications in the cinema market. The reason for this gap can be related with the type of data commonly used in the literature. One particularity of the cinema market is the availability of aggregated data bases (e.g. www.boxofficemojo.com). Despite the fact they are very useful for analyzing the evolution of box office performance, they do not provide information about the behavior of the spectators behind these aggregated results. Hence, instead of using these available databases, the study of the influence of superstars over individual spectators requires the generation of specific data. For this purpose, (a) some actors and actresses are carefully selected and (b) a questionnaire has been designed to measure superstars' persuasion.

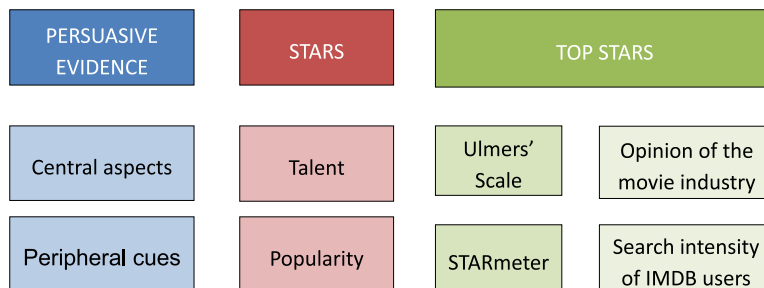


Figure 1: Superstar influence

2.1. Picking up actors and actress

There are different rankings of superstars depending on which of two facets of stardom, talent or popularity, is considered. On the one hand, the most important ranking of superstar talent is the Ulmer's Scale (www.ulmerscale.com). This scale analyzes the evolution of more than 1,400 stars based on the results of a global survey of movie











											
1. Brad Pitt	2. Christian Bale	3. George Clooney	4. Gerald Butler	5. Johnny Depp	6. Kristen Stewart						
Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm
3	5	-	7	5	-	-	8	2	1	-	2
											
7. Leonardo DiCaprio	8. Megan Fox	9. Nicholas Cage	10. Reese Witherspoon	11. Robert Downey	12. Robert Pattinson						
Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm
9	4	-	9	8	-	7	-	-	6	-	3
											
13. Russell Crowe	14. Tom Hanks	15. Will Ferrell	16. Will Smith	17. Zoe Saldana							
Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm	Ulmer	STARm		
10	-	4	-	6	-	1	-	-	10		

Figure 2: 2010 Superstar ranking (Ulmer's Scale/STARmeter)

buyers, sellers, sales agents, producers, company chiefs and financiers. On the other hand, the most relevant ranking of popularity is the STARMeter. It is elaborated by the leading cinema portal Internet Movie Data Base (IMDB) (www.imdb.com). STARMeter is based upon the actual search behavior of more than 100 million monthly unique users of IMDB. The results are an aggregate of weekly STARMeter rankings. The Ulmer's Scale is elaborated departing from the opinion of the industry, while the STARMeter takes into account users search intensity in IMDB (see Figure 1). This paper deals with the 10 actors or actresses placed at the top of both scales. It allows us to consider the best stars of the world according to the two main facets of stardom, talent and popularity.

Figure 2 presents the ten most talented stars according to the Ulmer's Scale 2010 and the ten most popular stars in STARMeter 2010. It also shows the position of each actor or actress in both scales. A dash indicates that this actor or actress is not placed at the 10 top of this particular ranking. Notice that there are three actors (namely Brad Pitt, Johnny Depp and Leonardo di Caprio) that are placed at the top of both rankings.

2.2. Designing a questionnaire to measure superstars' persuasion

The questionnaire created in this research takes into account several dimensions of superstars persuasion, both about variables related with central aspects of their persua-

sive capacity and with peripheral cues. Thus, the persuasion capacity of the stars can be the result of what is known about the actor (central aspects) or a consequence of the attitude towards the stars or emotional responses associated with the stars (peripheral cues). In particular, the questionnaire includes the following information:

- Knowledge about the stars. The survey respondents are asked to what extent they know about different aspects of the actors and actresses. It was measured by a 5-point scale adapted from (Hoyer and Brown, 1990).
- Attitude towards the stars. The survey respondents must answer to what extent actors and actresses adopt certain attitudes. It was operationalized by means of a 5-point Likert scale and based on 11 items developed by (Christensen, 2004). However, after a pretest of the original questionnaire, 2 items were removed for being a synonym and antonym of others already included, hence, leading to just considering 9 of the items in the final version of the questionnaire.
- Emotional responses associated with stars. In this case, survey respondents are asked to what extent they associate the actors and actresses with certain emotions. The inventory of emotions proposed for the hierarchical consumer emotions model of (Laros and Steenkamp, 2005) was used. This model orders emotions in three levels: superordinate level (positive or negative valence of the affect); intermediate level (8 basic emotions) and subordinate level (42 specific emotions). This paper deals with the intermediate level. This level is integrated by 8 basic emotions, four of a positive nature and four negative. The individuals have to initially value in which grade they experience each of these basic emotions in relation with each of the stars considered. The emotions were measured on a 5-point Likert scale where 1 means ‘I do not feel this emotion at all’ and 5 ‘I strongly feel this emotion’.
- Intention of seeing a movie. We use this variable as a measure of spectators response to the presence of superstars in the cast of a movie, since it allows us to gauge superstars’ persuasion. We asked about the intention of seeing a movie knowing that a certain actor or actress is the main star of the movie. Hence, it includes the measure of intention by a single-item scale. This alternative is very frequent in

shopping behavior studies (Christensen, 2004). Again, responses were rated on a 5-point Likert scale.

Table 1: Items of the questionnaire

Knowledge	Attitude	Emotions
His/her name rings a bell	Dynamic	Anger
I identify his/her physical appearance	Modern	Fear
I know some features of his/her personal life	Extroverted	Sadness
I have seen some of his/her movies	Simple	Shame
	Egocentric	Contentment
	Solidarity	Happiness
	Conservative	Love
	Prestigious	Pride
	Attractive	

After discussing and revising exhaustively the questionnaire, 21 items were included in the final version as measures of central aspects and peripheral cues of superstars' persuasive capacity in addition to the intention of seeing a movie (see Table 1).

The present study is based on a random sample of 320 individuals selected from a city in the north of Spain. All the individuals are aged between 20 and 34. They were chosen in this range because it is the most important segment of the cinema market (Collins and Hand, 2005). Besides filling in the questionnaire, they were asked their gender, age and frequency of cinema attendance. This additional information allows to classify the survey respondents into different groups taking into account both sociodemographic aspects and cinema habits. The main sample properties are summarized in Table 2.

3. Ordinal regression/classification statement and methods

This section deals with the problem statement and the machine learning techniques to tackle it.

Table 2: Sample profile

CINEMA ATTENDANCE	RATE
Less than five times per year	28.7%
Between five and six times per year	35.0%
Once per month	20.0%
Two or three times per month	12.5%
Once per week or more	3.8%
Total	100%
AGE	RATE
Mean	22.5
s.d.	2.6
GENDER	RATE
Male	48.8%
Female	51.2%
Total	100%

3.1. Ordinal regression/classification statement

The knowledge, the attitudes and the emotions can be considered as the inputs of the machine learning techniques, commonly called features, whereas the intention to see a movie can be taken as the output or target variable, typically called class or category. Despite the fact that all input variables (features) take ordinal values, no special treatment has been applied to them. However, the fact that the values of the class exhibit an order may provide useful information that can potentially improve the predictive performance of the models (Huhn and Hullermeier, 2008). This kind of problem is called ordinal regression or ordinal classification among researchers of the field. There are many problems of this kind recently solved under this paradigm, for instance, for text classification (Baccianella et al., 2013) where documents are classified in an ordinal scale, for predicting fertility rate (Piles et al., 2013) or for analyzing the migrants remitting patterns (Campoy-Muñoz et al., 2014). On the one hand, this problem can be considered

as a classification task because the values of the class are in fact labels, although they are commonly represented by numbers. For instance, if the intention is established to range from 1 to 5, then this is the same as, for instance, if it is established to take the values “very low intention”, “low intention”, “medium intention”, “high intention” and “very high intention”, which are discrete labels. On the other hand, the problem can be regarded as a regression task because there exists an order relation among the values of the class. Despite this fact, there is not a metric in the class values, since, for instance, an intention of 4 does not mean the double intention of 2 as happens if it is considered as a pure regression task. In other words, the examples are labeled by ranks, which exhibit an ordering among the different classes. In contrast to regression problems, these ranks are finite, and the metric distances between the ranks are not defined. These ranks are also different from the labels of multiple classes in classification problems due to the existence of the ordering information. Hence, the task is different either from classification or regression, although it shares properties with both. For this reason, researchers have been exploring specific methods for this task. An experimental study (Huhn and Hullermeier, 2008) explores to what extent existing techniques and learning algorithms for ordinal classification/regression are able to exploit order information, and which properties of these techniques are important in this regard. The main conclusions extracted from the paper are that learning techniques specifically designed for ordinal classification/regression are indeed able to exploit order information about classes and that the less flexible the learner is, the more it benefits from the ordinal structure of the data. This is the reason why this paper focused on treating the problem as it actually is, that is, as an ordinal classification/regression task, thus exploring techniques that have been specifically designed for this purpose.

The general framework of ordinal classification/regression can be formally stated as a finite set of ordered classes $C_1 < C_2 < \dots < C_q$ and a training set $D = \{(x_i, y_i) \in X \times Y \mid i = 1, \dots, n\}$ of size n , where $X \subseteq \mathbb{R}^k$ are the features and $Y = \{C_1, C_2, \dots, C_q\}$ is the set of classes, whose goal is to induce a hypothesis $h : X \rightarrow Y$ able to predict the class for any $x \in X$ according to the best fit possible. Notice that this task consists of estimating the probability $P(y/x)$ for any $y \in Y$.

3.2. Methods for ordinal regression/classification

Before continuing to discuss methods that deal with ordinal data, let us clarify that the goal is to perform a study of the imps of certain variables on the intention of seeing a movie rather than generating a ranking of the survey respondents from their ordinal responses. This fact allows to discard recommendation systems that generate a ranking from ordinal data such as the works reported in (Har-Peled et al., 2002) or (Herbrich et al., 2000).

There are several specific approaches in the literature that treat ordinal data providing an ordinal response rather than a ranking. At first glance, one can convert an ordinal scale into real values and then solve the problem as a standard regression problem (Kramer et al., 2000). But the model generated may not be reliable, since the metric distances between ordinal values are commonly unknown (the question is if the difference between saying “very low intention” and “low intention” is the same as between saying “high intention” and “very high intention”). The answer is that it is not possible to assure this assumption. Binary decomposition and mostly threshold approaches are the most widely applied for ordinal classification/regression (Campoy-Muñoz et al., 2014; Gutiérrez et al., 2012). Let us discuss both kinds of approaches.

3.2.1. Binary decomposition approaches

Decomposition approaches are based on decomposing the original problem in binary classification tasks, of course taking into account in some way the ordinal property of the problem. This was the idea proposed by (Frank and Hall, 2001), called Ordered Partitions (OP). They created a set of $q - 1$ chained binary classification problems where the i problem consists of separating the classes from 1 to i from the classes from $i + 1$ to q , $i = 1, \dots, q - 1$. Hence, the problem i estimates in fact the probability $P(y > i/x)$, that is, the probability of an example of having a class greater than i in the ordinal scale. Consequently, this method requires a base learner that can output class probability estimates. There are several ways of providing probability estimations, for instance, using logistic regression (Lin et al., 2008) or using the approach reported in (Platt, 1999).

$$P(y = 1|x, w) = \frac{1}{1 + \exp^{-y(w^T x + b)}} \quad (1)$$

minimizing the log-likelihood

$$\min_{w,b} \sum_{i=1}^n \log(1 + \exp^{-y_i(w^T x_i + b)}) \quad (2)$$

Once the models that estimate $P(y > i/x)$ are built, a composition approach is needed to estimate the probability of an example of being of class i , that is $P(y = i/x)$. In this direction, the proposal is quite straightforward, since it estimates $P(y = 1/x)$ by $1 - P(y > 1/x)$ for the lowest class $i = 1$. In the case of intermediate classes $1 < i < q$, $P(y = i/x)$ is estimated computing $P(y > i - 1/x) - P(y > i/x)$. Finally, the probability of the highest class $i = q$, that is, $P(y = q/x)$ will be $P(y > q - 1/x)$. The main problem of this appealing approach is that it does not guarantee that it will obtain an optimal solution. This is because it only takes into account the order in the frontiers of the classes and it does not optimize the global order.

3.2.2. Threshold approaches

Threshold approaches model ordinal problems under a regression perspective. Hence, they assume that some underlying real-value exists in the class although unobservable. Consequently, they firstly obtain a model that predicts a real value, but trying to dismiss the effect of making such assumption and they secondly obtain a set of thresholds to model the order of the classes.

Perhaps, the most classical and popular threshold approach is the Ordinal Logistic Regression (OLR) (Kleinbaum and Klein, 2010; McCullagh, 1980), also called Proportional Odds Model (POM), which is designed under a statistical perspective. It assumes a distribution of the unobservable class variable and a stochastic order of the input space X . The method takes cumulative probabilities up to a threshold, then making the whole range of ordinal classes binary at that threshold. If p_1, \dots, p_q are the probabilities of the respective classes C_1, \dots, C_q , then a cumulative probability of the class being less than or equal to i is $P(y \leq i/x) = p_1 + \dots + p_i$. Then a cumulative logit for class C_i is defined as

$$\ell_i = \log \left(\frac{P(y \leq i/x)}{P(y > i/x)} \right) = \log \left(\frac{P(y \leq i/x)}{1 - P(y \leq i/x)} \right) = \log \left(\frac{p_1 + \dots + p_i}{p_{i+1} + \dots + p_q} \right) \quad (3)$$

which defines the log-odds of two cumulative probabilities, namely, of being less than class i and of being greater than class i . This log-odds evaluates to what extent the

prediction is equal to or lower than i against being higher than i . The method obtains a set of hyperplanes satisfying the following equations

$$\ell_i = b_i + w \cdot x \quad (4)$$

Hence, a set of $q - 1$ parallel hyperplanes is obtained since the slope coincides for all the classes i and they differ in a distance b_i . The parameters are estimated using the maximum likelihood. As i increases b_i also increases since the ℓ_i also increases. Then, the probabilities of each class C_i are

$$P(y = i/x) = P(y \leq i/x) - P(y \leq i - 1/x) \quad (5)$$

where $P(y \leq i/x)$ in terms of w and b_i is

$$P(y \leq i/x) = \frac{\exp^{b_i - wx}}{1 + \exp^{b_i - wx}} \quad (6)$$

A later threshold approach (Shashua and Levin, 2002) applies the large margin principle used in support vector methods (Vapnik, 1995). It finds $q - 1$ thresholds that represent $q - 1$ parallel hyperplanes whose direction is w for the q ordered classes, but the ordinal inequalities are not taken into account, so the method may provide a solution with disordered thresholds with regard to the order of the classes. A more recent work (Chu and Keerthi, 2007) takes the idea of (Shashua and Levin, 2002), but including the thresholds in the formulation of the problem, hence solving the drawback of obtaining disordered thresholds. Two ways of including the order of the classes are proposed (Chu and Keerthi, 2007). The first, called Explicit Support Vector Regression (SVREX), formulates the problem exactly as (Shashua and Levin, 2002), but explicitly including certain constraints in the problem formulation that enforce the inequalities on the thresholds. The second, called Implicit Support Vector Regression (SVORIM) formulates the problem in a different way including the constraints in an implicit way. Let us now go into both approaches in detail. Concerning SVOREX, it finds $q - 1$ thresholds that represent $q - 1$ parallel hyperplanes whose direction is w for the q ordered classes accordingly. This means that $q - 1$ margins are allowed whose sum is maximized (or errors minimized). Then, taking the samples of two consecutive classes j and $j + 1$, (i) each example x_i^j of the class j should have a function value that is less than the lower margin $b_j - 1$; otherwise $\xi_i^j = \langle w \cdot \phi(x_i^j) \rangle - (b_j - 1)$ is the error for this example and

similarly (ii) each example x_i^{j+1} of the class $j + 1$ should have a function value that is greater than the upper margin $b_j + 1$; otherwise $\xi_i^{*j+1} = (b_j + 1) - \langle w \cdot \phi(x_i^{j+1}) \rangle$ is the error for this example. Hence, the optimization problem that they solve is formulated as follows

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \langle w \cdot w \rangle + c \sum_{j=1}^{q-1} \left(\sum_{i=1}^{n^j} \xi_i^j + \sum_{i=1}^{n^{j+1}} \xi_i^{*j+1} \right) \quad (7)$$

subject to

$$\langle w \cdot \phi(x_i^j) \rangle - b_j \leq -1 + \xi_i^j \quad \xi_i^j \geq 0 \quad \text{for } i = 1, \dots, n^j \quad (8)$$

$$\langle w \cdot \phi(x_i^{j+1}) \rangle - b_j \geq +1 - \xi_i^{*j+1} \quad \xi_i^{*j+1} \geq 0 \quad \text{for } i = 1, \dots, n^{j+1} \quad (9)$$

where $j = 1, \dots, q - 1$, $c > 0$ is a regularity parameter for avoiding overfitting that must be optimized, $n = \sum_{j=1}^q n^j$ is the number of examples and $\phi(\cdot)$ is the kernel function that transforms the examples into another Hilbert space where linear hyperplanes will be obtained rather than other complex functions. The main problem of this approach is that it does not include the ordinal inequalities on the thresholds, that is, it is possible that the inequalities $b_1 \leq b_2 \leq \dots \leq b_{q-1}$ do not hold, which may lead to undesirable disordered thresholds in some cases. Hence, the constraints $b_{j-1} \leq b_j$ for $j = 2, \dots, q - 1$ must be explicitly included. The optimization problem is solved via Lagrangian strategies. So, using Lagrangian multipliers $(\alpha_i^j, \alpha_i^{*j}, \mu^j)$ to cope with the inequalities and imposing the Kuhn-Tucker conditions, a convex quadratic programming problem is obtained

$$\max_{\alpha, \alpha^*, \mu} \sum_{j,i} (\alpha_i^j + \alpha_i^{*j}) - \frac{1}{2} \sum_{j,i} \sum_{j',i'} (\alpha_i^{*j} - \alpha_i^j) (\alpha_{i'}^{*j'} - \alpha_{i'}^{j'}) K(x_i^j, x_{i'}^{j'}) \quad (10)$$

subject to

$$0 \leq \alpha_i^j \leq c \quad \forall i, j \quad (11)$$

$$0 \leq \alpha_i^{*j+1} \leq c \quad \forall i, j \quad (12)$$

$$\sum_{i=1}^{n^j} \alpha_i^j + \mu^j = \sum_{i=1}^{n^{j+1}} \alpha_i^{*j+1} + \mu^{j+1} \quad \forall j \quad (13)$$

$$\mu^j \geq 0 \quad \forall j \quad (14)$$

Once the optimization problem is solved, the final decision function for a new example x is

$$f(x) = \langle w \cdot \phi(x) \rangle = \sum_{j,i} (\alpha_i^{*j} - \alpha_i^j) K(x_i^j, x) \quad (15)$$

and the predictive ordinal decision is

$$\arg \min_i \{i : f(x) < b_i\} \quad (16)$$

The particularity of SVORIM guarantees that the thresholds are properly ordered at the optimal solution. In fact, the ordinal inequality constraints on the thresholds are automatically satisfied, though there are no explicit constraints on these thresholds. In SVORIM, the errors ξ and ξ^* are redefined; so (i) the function values of all the examples from all the lower classes than j should be less than the lower margin $b_j - 1$; otherwise $\xi_{ki}^j = \langle w \cdot \phi(x_i^k) \rangle - (b_j - 1)$ is the error of x_i^k for the threshold b_j for all classes $k \leq j$ and similarly (ii) the function values of all the samples from all the upper classes than j should be greater than the upper margin $b_j + 1$; otherwise $\xi_{ki}^{*j} = (b_j + 1) - \langle w \cdot \phi(x_i^k) \rangle$ is the error of x_i^k for the threshold b_j for all classes $k > j$. Hence, taking into account this new definition of ξ and ξ^* , the formulation of the problem is transformed into

$$\min_{(w,b,\xi,\xi^*)} \frac{1}{2} \langle w \cdot w \rangle + c \sum_{j=1}^{q-1} \left(\sum_{k=1}^j \sum_{i=1}^{n^k} \xi_{ki}^j + \sum_{k=j+1}^q \sum_{i=1}^{n^k} \xi_{ki}^{*j} \right) \quad (17)$$

subject to

$$\langle w \cdot \phi(x_i^k) \rangle - b_j \leq -1 + \xi_{ki}^j \quad \xi_{ki}^j \geq 0 \quad \text{for } k = 1, \dots, j \quad \text{and } i = 1, \dots, n^k \quad (18)$$

$$\langle w \cdot \phi(x_i^k) \rangle - b_j \leq +1 - \xi_{ki}^{*j} \quad \xi_{ki}^{*j} \geq 0 \quad \text{for } k = j+1, \dots, q \quad \text{and } i = 1, \dots, n^k \quad (19)$$

An illustration of the idea of this algorithm is presented in Figure 3. In this example there are three ordered classes ($q = 3$), one per kind of circles. Then, two ($q - 1 = 2$) thresholds (b_1 and b_2) are obtained for the two hyperplanes of vertical direction ($w =$

(0, 1)). Several situations can be seen in this graph: a) The example A of class 1 (white circles or $y = 1$) has an error (ξ_{1i}^1) because it is located to the right of the left margin ($b_1 - 1$). Besides, it is misclassified since it is to the right of the threshold b_1 ; b) The example B of class 2 (light colored circles or $y = 2$) has an error (ξ_{2i}^{*1}) because it is placed to the left of the right margin ($b_1 + 1$). But, it is not misclassified since it is to the right of the threshold b_1 ; c) The example C of class 3 (dark colored circles or $y = 3$) has two errors, (ξ_{3i}^{*1}) because it is located to the left of the right margin $b_1 + 1$ and ξ_{3i}^{*2} for being placed to the left of the right margin $b_2 + 1$. Besides, it is misclassified twice, one for being to the left of threshold b_1 and the other for being to the left of threshold b_2 ; d) The example D of class 2 has an error ξ_{2i}^2 since it is placed to the right of the left margin $b_2 - 1$. Besides, it is misclassified because it is to the right of threshold b_2 .

Following the same process as before using Lagrangian multipliers (now using just α_i^j and α_i^{*j}) and Kuhn-Tucker constraints, the problem is converted into another convex quadratic programming problem

$$\max_{\alpha, \alpha^*} - \frac{1}{2} \sum_{k,i} \sum_{k',i'} \left(\sum_{j=1}^{k-1} \alpha_{ki}^{*j} - \sum_{j=k}^{q-1} \alpha_{ki}^j \right) \left(\sum_{j=1}^{k'-1} \alpha_{k'i'}^{*j} - \sum_{j=k'}^{q-1} \alpha_{k'i'}^j \right) K(x_i^k, x_{i'}^{k'}) + \dots \quad (20)$$

$$\dots + \sum_{k,i} \left(\sum_{j=1}^{k-1} \alpha_{ki}^{*j} + \sum_{j=k}^{q-1} \alpha_{ki}^j \right) \quad (21)$$

subject to

$$0 \leq \alpha_{ki}^j \leq c \quad \forall j \quad \text{and} \quad k \leq j \quad (22)$$

$$0 \leq \alpha_{ki}^{*j} \leq c \quad \forall j \quad \text{and} \quad k > j \quad (23)$$

$$\sum_{k=1}^j \sum_{i=1}^{n^k} \alpha_{ki}^j = \sum_{k=j+1}^q \sum_{i=1}^{n^k} \alpha_{ki}^{*j} \quad \forall j \quad (24)$$

Again, once the optimization problem is solved, the final decision function for a new example x is

$$f(x) = \langle w \cdot \phi(x) \rangle = \sum_{k,i} \left(\sum_{j=1}^{k-1} \alpha_{ki}^{*j} - \sum_{j=k}^{q-1} \alpha_{ki}^j \right) K(x_i^k, x) \quad (25)$$

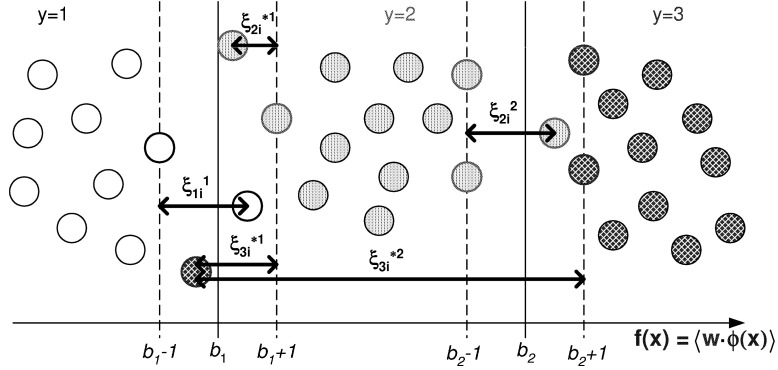


Figure 3: Graphical interpretation of ξ and ξ^*

and the predictive ordinal decision remains equal to

$$\arg \min_i \{i : f(x) < b_i\} \quad (26)$$

The size of the optimization problems of both SVOREX and SVORIM remains linear in the number of examples. Moreover, the optimization algorithms are easy to implement and scales efficiently as a quadratic function of the problem size. These are in fact two well-known and renowned methods among researchers for ordinal classification/regression that have shown good performance (Chu and Keerthi, 2007; Gutiérrez et al., 2012). But, since none of these works opt for either, just concluding that both are good, we opt for SVORIM because of the particularity that the ordinal inequality constraints on the thresholds are automatically satisfied at the optimal solution. From now on, for the sake of simplicity, SVORIM will be referred to as just SVOR.

4. Experiments

This section deals with the data description, details the methods taken to compare with the ordinal approaches and the experimental settings, and finally, the results are discussed.

4.1. Dataset description

We have analyzed the superstars' influence over spectators from two different points of view. On the one hand, we were interested in studying if the persuasion differs between

the male and female audience. Hence, the spectators' answers were split into two groups, one from males and the other from females.

Moreover, for each of these two segments, each actor and actress has been taken separately. Hence, a total of 4×17 different data sets were built. Each of these data sets has the responses of 320 individuals about the 21 items described in Table 1 as features and the intention of seeing a movie as the class.

4.2. Non ordinal classification/regression methods vs. ordinal ones

Several models have been built for each data set, since several methods have been taken for comparison. Since ordinal classification/regression is a task between classification and regression as discussed in Section 3.1, we take both classification and regression methods to compare with the ordinal approaches. Firstly, the median of the class is obtained as baseline classification method. It consists of calculating the median of the class, without taking into account the information provided by the features. Secondly, the classical least squares (LS) method for regression is also taken for comparison. Furthermore, since SVOR is one of the ordinal methods we take, we also choose the classification version, namely, Support Vector Machines (SVM) and the regression version, namely, Support Vector Regression (SVR). In both cases, it means relaxing the problem and ignoring the information that provides the order of the values of the class in case of SVM and assuming the distance between the values of consecutive classes to be equal in the case of SVR. The implementation taken for both SVM and SVR is that reported in (Chang and Lin, 2011). Finally, models using the ordinal approaches OLR, OP and SVOR are also built. The implementation considered for the base learner of OP is the one reported in (Chang and Lin, 2011) for estimating probabilities based on (Platt, 1999).

4.3. Experimental settings

Concerning the parameters of SVM, SVR and SVOR, both linear and gaussian kernels were taken.

In the case of linear kernel, the regularization parameter c was established performing a grid search over the values $c \in \{10^p, p \in \{-2, -1, 0, 1, 2\}\}$ optimizing the mean absolute error estimated by means of a balanced 2-fold cross validation repeated 3 times. In the case of gaussian kernel, both the regularization parameter c and the kernel parameter γ

were set in the same way, but γ taking values in $\{(10^p)/2, p \in \{-3, -2, -1, 0, 1, 2, 3\}\}$. The mean absolute error used as loss function to be optimized is also employed as evaluation metric for the results through a 5-fold cross validation repeated twice. This loss function and evaluation metric is defined as the average deviation of the prediction from the true target, that is, the ordinal scales are treated as consecutive integers. The expression is as follows:

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (27)$$

where y_i is the actual value and $\hat{y} = h(x_i)$ is the predicted value through h for instance x_i . Hence, for example, if the method predicts an intention of seeing a movie of 5, but the real intention is 3, then the error will be 2 units. This type of measure is preferable for ordinal data instead of the mean one-zero error that only counts the errors. In the previous example, the one-zero mean will be 1 unit. It is an error and the difference does not matter. The reason is quite obvious, since the former is able for instance to differentiate the case of predicting an intention of 4 from predicting an intention of 2 when the actual value is 5, whereas the latter is not because it considers both errors having equal importance. In this example, the mean absolute error is 1 unit if the model predicts 4 instead of 5 and it is 3 units if the model predicts 2 instead of 5, whereas the one-zero error is 1 unit in both cases. Obviously, for ordinal data predicting 4 instead of 5 is better than predicting 2 instead of 5. The mean absolute error is commonly called linear loss within researchers of the field, whereas the one-zero error is usually called multiclass loss.

4.4. Results and discussion

This section describes and discusses the results of the experiments performed. Tables 3 and 4 respectively show the linear loss for the two segments in which the answers were split when all knowledge about stars, attitude towards the stars and emotional responses associated with stars are considered as features. They include the performance of the classification methods Median and SVM, the regression approaches LS and SVR and the ordinal techniques ORL, OP and SVOR for the 17 stars. The subindex l and g in SVM, SVR and SVOR indicates the kernel taken. The stars are numbered according to Figure 2. Numbers in brackets indicate the position of each method according to its

performance. The last line is the averaged rank (A. r.) over all stars. Notice that the evaluation metric is a loss, hence, the lower the value, the better the performance.

Moreover, following the recommendations of (Demšar, 2006) a two-step comparison for each of the considered measures was performed. The first step consists of a Friedman test of the null hypothesis that states that all approaches perform equally. Then, in the case where this hypothesis is rejected, the Nemenyi test is performed to compare the methods in a pairwise way. Both tests are based on the rank average. The critical differences (CD) in the Nemenyi test depend on the number of datasets and learners compared. They are shown in the caption of each table for different significance levels.

In view of these results, it is clear that for the male public SVR and SVOR are statistically better at 95% than Median, LS, SVM and even than OLR whatever kernel is taken. However, in the case of female public, OP using Gauss kernel and SVOR are statistically better than them.

Despite the bad performance of the Median, LS and SVM and that there are no significant differences between them, their results provide useful information. Let us remember that (a) Median considers the order of the target variable but it does not take into account either the knowledge about the stars or the attitude towards them, or the emotional responses associated with them; (b) LS takes into account both kinds of information, although, as a regressor, it assumes a metric in the values of the target variable and (c) SVM considers the knowledge about the stars, the attitude towards them and the emotional responses associated with them, but it ignores the order of the target variable. Being aware of this fact, we can conclude that (i) both kinds of information help to obtain better models since LS performs better than Median and SVM (only with linear kernel in case of female public) and (ii) ignoring the order information of the target variable is slightly less serious than using neither knowledge about the stars, nor the attitude towards them, nor the emotional responses associated with them, since the performance of SVM is slightly better than LS. In this sense, the experimental results support the theoretical hypothesis that knowledge about the stars, the attitude towards them and the emotional responses associated with them manage to explain the intention of watching a movie. Hence, superstar persuasion is experimentally shown to be the result of central aspects related with the superstar and the product of peripheral cues.

Curiously, the regressor SVR performs significantly better than the ordinal approach OLR. So, the robustness of SVR even considering a metric in the target variable prevails over OLR that ignores that metric. SVOR enjoys this robustness of SVR, but like OLR providing parallel hyperplanes for separating the ordinal classes.

Comparing SVR and SVOR, it happens that the latter is slightly better than the former, although there are not significant differences between them. On the one hand, the differences in the performance between SVM and SVR confirm us that the problem is closer to being a regression task more than a classification task. Hence, it is confirmed again that the target variable, that is, the intention of watching a movie, exhibits an order. On the other hand, SVOR presents a slightly better performance than SVR. This fact allows to conclude that taking into account a metric between the values of the target variable adds noise, since SVOR does not take into account a metric between the value of the classes as SVR does. Therefore, the experiments confirm the theoretical statement of considering the problem as an ordinal task. This fact is quite interesting, since not in vain, intention is commonly used as a response variable in many shopping behavior models. However, these results show that the predictions miss important nuances when this variable is treated as a metric. The results also point out some differences between the male and female public in this respect, since there exist many more differences between SVR and SVOR in female spectators than in male spectators, although they are not significative. In fact, for the female public the ordinal approach OP with gaussian kernel falls in between SVR and SVOR. This different behavior makes us think that female spectators focus more than male spectators on the semantic meaning of the 5-point scale of the target variable rather than on the number of the scale when they mark the answer. To the best of our knowledge, gender has not been considered to evaluate the convenience and implications of using different types of scales, so this fact may mark an interesting path for future research.

Besides, the best results are obtained using gaussian kernel rather than linear kernel. Hence, it seems that the relationship between the persuasion items (knowledge about the stars, the attitude towards them and the emotional responses associated with them) and the intention of watching a movie is not linear. However, due to the fact that the differences are not significative we can state that the relationship, although not linear,

is close to being linear.

Let us now analyze now the knowledge of the stars and attitude and emotions towards them on their own. Tables 5 and 6 respectively show the linear loss for the 17 actors/actresses and for the two segments in which the answers were split, considering different configurations of the components knowledge about stars, attitude towards the stars and emotional responses associated with stars. The experiments with SVOR and gaussian kernel were taken into account, since this method has shown in general the best results as discussed above.

The results allow to conclude that all the components, namely, knowledge about stars and attitude and emotions toward them are relevant in order to predict the intention of seeing a movie in which they appear, since the best results are obtained in this case both for male and female public. Unfortunately, this fact does not happen for some stars. In these cases attitude plays a key role. Considering just the knowledge of the stars or just the emotions towards them significantly produces the worst results. However, just taking the items about attitude towards stars is not bad at all, hence, it seems that attitude towards stars is key information to gather. In the case of male public adding to attitude items knowledge about the stars or emotions towards them improves the performance, especially emotions. But the improvement is not so important as with the female audience. In this case, both attitude and emotions really enhance the performance. Again, these results show that the routes of superstars persuasion are different for the male and female public. This conclusion is very relevant for the cinema industry because it signs the adequate type of information that triggers male (knowledge-attitude) and female (knowledge-attitude-emotion) public. It has also important consequences for research in the field of stardom. It points out the relevance of the gender as a variable to explain movie-goers behavior.

5. Conclusions and future work

This paper studies the impact of considering superstar persuasion to explain the intention of watching a movie. Here, persuasion is represented by three kinds of features, namely, knowledge about stars, attitude towards them and emotional responses associated with them. The problem is stated as an ordinal classification/regression task rather

than a traditional classification or regression task. This means to exploit the property of the target variable consisting of exhibiting an order. The experiments carried out confirm that this formulation is a success, since the methods employed beat others that formulate the problem either as a traditional classification or as a regression tasks. Moreover, the above-mentioned features are shown to be useful for inducing accurate models, that is, both talent (central aspects) and popularity (peripheral cues) allow to quantify, to a certain extent, the power of superstars in the movie industry. Moreover, attitude towards actors is key information. However, in the case of female audience both emotions and attitude play a relevant role. This study helps to guide the task of creating and managing a superstar when they have to be chosen for an endorsement. Finally, this work allows to overcome the paucity of research both in micro analysis of the cinema market and in the application of machine learning methods to solve marketing problems.

An immediate proposal for future research is to analyze more in detail which items taken for knowledge about stars, for attitude towards the stars and for emotional responses associated with stars are in fact relevant and useful, non redundant and not noisy in order to improve the questionnaires in this sense. For this purpose, we plan to analyze this issue through feature selection and extraction techniques. Also, consumer and industrial behavior is growing in importance as an area for the applications of expert and intelligence systems. In this sense, other possible directions for future research will include the following. It could be very useful to analyze whether the outcomes of the approach vary when on line surveys vs. personal surveys are carried out. On the other hand, we could compare our conclusions with those resulting from using social network content instead of data from questionnaires. Another possible direction for future research would include attempting to explain how intentions lead to actions. The results of these comparisons could be interesting for many other applications of expert and intelligence systems in marketing settings.

Acknowledgements

This research has been partially supported by the Spanish Ministerio de Economía y Competitividad, grant TIN2011-23558.

References

- Abernethy, J., Evgeniou, T., Toubia, O., Vert, J.P., 2008. Eliciting consumer preferences using robust adaptive choice questionnaires. *Knowledge and Data Engineering, IEEE Transactions on* 20, 145–155.
- Adler, M., 2006. *Stardom and talent*, 1 ed.. Elsevier. volume 1. chapter 25, pp. 895–906. URL: <http://EconPapers.repec.org/RePEc:eee:artchp:1-25>.
- Albert, S., 1999. Movie stars and the distribution of financially successful films in the motion picture industry. *Journal of Cultural Economics* 23, 325–329. URL: <http://dx.doi.org/10.1023/A%3A1007584017128>, doi:10.1023/A:1007584017128.
- Baccianella, S., Esuli, A., Sebastiani, F., 2013. Using micro-documents for feature selection: The case of ordinal text classification. *Expert Systems with Applications* 40, 4687 – 4696.
- Cacioppo, J.T., Berntson, G.G., Petty, R.E., 1997. Persuasion. *Encyclopedia of human biology* 5, 679 – 690.
- Campoy-Muñoz, P., Gutiérrez, P.A., Hervás-Martínez, C., 2014. Addressing remitting behavior using an ordinal classification approach. *Expert Systems with Applications* 41, 4752 – 4761.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27.
- Cheung, K.W., Kwok, J.T., Law, M.H., Tsui, K.C., 2003. Mining customer product ratings for personalized marketing. *Decision Support Systems* 35, 231 – 243.
- Christensen, S., 2004. Measuring consumer reactions to sponsoring partnerships based upon emotional and attitudinal responses. *Center for Marketing Communication* .
- Chu, W., Keerthi, S., 2007. Support vector ordinal regression. *Neural Computation* 19, 792–815.
- Collins, A., Hand, C., 2005. Analyzing moviegoing demand: an individual-level cross-sectional approach. *Managerial and Decision Economics* 26, 319–330.
- Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7, 1–30.
- Eliashberg, J., Elberse, A., Leenders, M.A.A.M., 2006. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science* 25, 638–661.
- Filimon, N., López-Sintas, J., Padró-Reig, C., 2011. A test of Rosens and Adlers theories of superstars. *Journal of Cultural Economics* 35, 137–161.
- Frank, E., Hall, M., 2001. A simple approach to ordinal classification, in: Raedt, L., Flach, P. (Eds.), *Machine Learning: ECML 2001*. Springer Berlin Heidelberg. volume 2167 of *Lecture Notes in Computer Science*, pp. 145–156.
- Ginsburgh, V., Weyers, S., 1999. On the perceived quality of movies. *Journal of Cultural Economics* 23, 269–283.
- Gutiérrez, P.A., Pérez-Ortiz, M., Fernández-Navarro, F., Sánchez-Monedero, J., Hervás-Martínez, C., 2012. An experimental study of different ordinal regression methods and measures, in: *HAIS (2)*, pp. 296–307.
- Hansen, F., i København. HHK. Institut for Afsætningsøkonomi. AØ, H., of marketing, C.B.S.C.D., 1997. Quantifying Creative Contributions: Advertising Pretesting’s New Generation. Working paper

- (Copenhagen Business School. Department of Marketing), Copenhagen Business School, Department of Marketing.
- Har-Peled, S., Roth, D., Zimak, D., 2002. Constraint classification: A new approach to multiclass classification, in: *Proceedings of the 13th International Conference on Algorithmic Learning Theory*, Springer-Verlag, London, UK, UK. pp. 365–379.
- Herbrich, R., Graepel, T., Obermayer, K., 2000. Large margin rank boundaries for ordinal regression, in: Smola, A., Bartlett, P., Schölkopf, B., Schuurmans, D. (Eds.), *Advances in Large Margin Classifiers*, MIT Press, Cambridge, MA. pp. 115–132.
- Hoyer, W.D., Brown, S.P., 1990. Effects of brand awareness on choice for a common, repeat-purchase product. *Journal of Consumer Research* 17, 141–148.
- Huhn, J.C., Hullermeier, E., 2008. Is an ordinal class structure useful in classifier learning? *International Journal of Data Mining, Modelling and Management* 1, 45–67.
- Jedidi, K., Krider, R., Weinberg, C., 1998. Clustering at the movies. *Marketing Letters* 9, 393–405.
- Kleinbaum, D., Klein, M., 2010. Ordinal logistic regression, in: *Logistic Regression*. Springer New York. *Statistics for Biology and Health*, pp. 463–488.
- Kramer, S., Widmer, G., Pfahringer, B., Groeve, M., 2000. Prediction of ordinal classes using regression trees, in: Ra, Z., Ohsuga, S. (Eds.), *Foundations of Intelligent Systems*. Springer Berlin Heidelberg. volume 1932 of *Lecture Notes in Computer Science*, pp. 426–434.
- Laros, F.J., Steenkamp, J.B.E., 2005. Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research* 58, 1437 – 1445.
- Lin, C.J., Weng, R.C., Keerthi, S.S., 2008. Trust region newton method for logistic regression. *J. Mach. Learn. Res.* 9, 627–650.
- Lord, K.R., Lee, M.S., Sauer, P.L., 1995. The combined influence hypothesis: Central and peripheral antecedents of attitude toward the ad. *Journal of Advertising* 24, 73–85.
- Marshall, P., Dockendorff, M., Ibez, S., 2013. A forecasting system for movie attendance. *Journal of Business Research* 66, 1800 – 1806.
- McCullagh, P., 1980. Regression models for ordinal data. *J. Roy. Statist. Soc. B* 42, 109–142.
- Nelson, R., Glotfelty, R., 2012. Movie stars and box office revenues: an empirical analysis. *Journal of Cultural Economics* 36, 141–166.
- Piles, M., Díez, J., del Coz, J., Montañés, E., Quevedo, J., Ramon, J., Rafel, O., López-Béjar, M., Tusell, L., 2013. Predicting fertility from seminal traits: Performance of several parametric and non-parametric procedures. *Livestock Science* , –.
- Platt, J.C., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: *Advances in Large Margin Classifiers*, MIT Press. pp. 61–74.
- Ravid, S.A., 1999. Information, blockbusters, and stars: A study of the film industry. *The Journal of Business* 72, 463–92.
- Rosen, S., 1981. The economics of superstars. *American Economic Review* 71, 845–58.
- Sedgwick, J., Pokorny, M., 2012. Film consumer decision-making: The philadelphia story, 193536. *Journal of Consumer Culture* 12, 323–346. URL: <http://joc.sagepub.com/content/12/3/323.abstract>,

arXiv:<http://joc.sagepub.com/content/12/3/323.full.pdf+html>.

Shashua, A., Levin, A., 2002. Ranking with large margin principle: Two approaches, in: Becker, S., Thrun, S., Obermayer, K. (Eds.), NIPS, MIT Press. pp. 937–944.

Suárez-Vázquez, A., 2011. Critic power or star power? the influence of hallmarks of quality of motion pictures: an experimental approach. *Journal of Cultural Economics* 35, 119–135.

Vapnik, V.N., 1995. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.

Table 3: Linear loss for male survey respondents when all knowledge about stars, attitude towards the stars and emotional responses associated with stars are considered as features. The critical differences for the Neményi test are: $CD_{p=0.05} = 3.66$ and $CD_{p=0.1} = 3.38$

	Classification					Regression					Ordinal										
	Median	SVM _l	SVM _g	LS	SVR _l	SVR _g	OLR	OP _l	OP _g	SVOR _l	SVOR _g	Median	SVM _l	SVM _g	LS	SVR _l	SVR _g	OLR	OP _l	OP _g	SVOR _l
1	1.24 (10)	1.31 (11)	1.21 (8)	1.13 (4.5)	1.07 (1)	1.10 (3)	1.21 (8)	1.20 (6)	1.21 (8)	1.08 (2)	1.13 (4.5)										
2	1.18 (11)	0.92 (8)	0.88 (6.5)	0.99 (10)	0.82 (3.5)	0.81 (2)	0.97 (9)	0.88 (6.5)	0.73 (1)	0.83 (5)	0.82 (3.5)										
3	1.01 (4.5)	1.15 (11)	1.09 (10)	1.03 (6.5)	0.97 (1)	0.99 (2)	1.06 (9)	1.00 (3)	1.03 (6.5)	1.04 (8)	1.01 (4.5)										
4	1.21 (11)	0.85 (7.5)	0.87 (10)	0.85 (7.5)	0.81 (4.5)	0.82 (6)	0.80 (3)	0.86 (9)	0.81 (4.5)	0.72 (1)	0.77 (2)										
5	1.02 (10.5)	0.98 (8.5)	1.02 (10.5)	0.96 (7)	0.92 (3.5)	0.92 (3.5)	0.93 (5.5)	0.89 (2)	0.98 (8.5)	0.93 (5.5)	0.88 (1)										
6	0.88 (8)	0.89 (10)	0.86 (6)	0.89 (10)	0.79 (2)	0.80 (4)	0.89 (10)	0.87 (7)	0.80 (4)	0.80 (4)	0.77 (1)										
7	1.23 (6)	1.48 (10)	1.30 (7.5)	1.18 (5)	1.08 (1)	1.10 (2)	1.34 (9)	1.51 (11)	1.30 (7.5)	1.16 (4)	1.14 (3)										
8	1.37 (11)	1.17 (7)	1.20 (9)	1.16 (6)	1.06 (3)	1.07 (4)	1.20 (9)	1.12 (5)	1.20 (9)	0.97 (1)	1.03 (2)										
9	1.12 (11)	1.03 (6)	1.03 (6)	1.05 (8.5)	0.99 (2)	1.00 (3.5)	1.07 (10)	1.05 (8.5)	0.98 (1)	1.03 (6)	1.00 (3.5)										
10	1.01 (10.5)	1.01 (10.5)	0.93 (7.5)	0.97 (9)	0.84 (3.5)	0.84 (3.5)	0.93 (7.5)	0.87 (6)	0.86 (5)	0.76 (1)	0.80 (2)										
11	1.10 (11)	1.04 (9)	1.07 (10)	0.98 (7)	0.87 (5)	0.86 (4)	1.00 (8)	0.94 (6)	0.83 (3)	0.82 (1.5)	0.82 (1.5)										
12	0.88 (8.5)	0.88 (8.5)	0.84 (5)	0.93 (11)	0.83 (3.5)	0.82 (2)	0.86 (6)	0.87 (7)	0.79 (1)	0.89 (10)	0.83 (3.5)										
13	1.23 (11)	1.16 (9)	1.15 (7)	1.10 (5)	0.99 (2)	0.99 (2)	1.16 (9)	1.16 (9)	1.13 (6)	1.00 (4)	0.99 (2)										
14	1.21 (8)	1.28 (11)	1.26 (10)	1.11 (5)	1.05 (2)	1.06 (3.5)	1.25 (9)	1.18 (6.5)	1.18 (6.5)	1.06 (3.5)	1.01 (1)										
15	0.92 (7)	0.92 (7)	0.94 (9)	0.99 (10)	0.85 (3)	0.88 (5)	1.02 (11)	0.82 (1)	0.84 (2)	0.86 (4)	0.92 (7)										
16	0.87 (11)	0.81 (8)	0.83 (9)	0.85 (10)	0.75 (5.5)	0.73 (3.5)	0.78 (7)	0.73 (3.5)	0.75 (5.5)	0.70 (2)	0.68 (1)										
17	0.82 (9)	0.74 (7.5)	0.73 (5.5)	0.83 (10)	0.72 (4)	0.73 (5.5)	0.84 (11)	0.65 (1)	0.74 (7.5)	0.68 (3)	0.66 (2)										
A. r.	(9.35)	(8.79)	(8.03)	(7.76)	(2.94)	(3.47)	(8.29)	(5.76)	(5.09)	(3.85)	(2.65)										

Table 4: Linear loss for female survey respondents when all knowledge about stars, attitude towards the stars and emotional responses associated with stars are considered as features. The critical differences for the Nemenyi test are: $CD_{p=0.05} = 3.66$ and $CD_{p=0.1} = 3.38$

	Classification										Regression										Ordinal			
	Median	SVM _t	SVM _g	LS	SVR _t	SVR _g	OLR	OP _t	OP _g	SVOR _t	SVOR _g	Median	SVM _t	SVM _g	LS	SVR _t	SVR _g	OLR	OP _t	OP _g		SVOR _t	SVOR _g	
1	0.99 (7)	1.09 (11)	0.97 (6)	1.02 (8)	0.91 (3)	0.90 (2)	1.04 (10)	1.03 (9)	0.89 (1)	0.89 (1)	0.92 (4)	1.03 (9)	1.03 (9)	0.89 (1)	0.89 (1)	0.95 (5)	0.95 (5)	1.04 (10)	1.03 (9)	0.89 (1)	0.89 (1)	0.95 (5)	0.92 (4)	
2	1.03 (11)	0.97 (10)	0.89 (6.5)	0.92 (8)	0.81 (2.5)	0.83 (4)	0.93 (9)	0.89 (6.5)	0.78 (1)	0.81 (2.5)	0.84 (5)	0.93 (9)	0.89 (6.5)	0.78 (1)	0.81 (2.5)	0.81 (2.5)	0.84 (5)	0.93 (9)	0.89 (6.5)	0.78 (1)	0.81 (2.5)	0.81 (2.5)	0.84 (5)	
3	1.07 (10)	0.98 (8)	0.93 (5)	0.95 (7)	0.92 (3)	0.92 (3)	1.06 (9)	1.09 (11)	0.92 (3)	0.94 (6)	0.87 (1)	1.06 (9)	1.09 (11)	0.92 (3)	0.94 (6)	0.94 (6)	0.87 (1)	1.06 (9)	1.09 (11)	0.92 (3)	0.94 (6)	0.94 (6)	0.87 (1)	
4	1.11 (11)	0.84 (9)	0.92 (10)	0.79 (7)	0.73 (3.5)	0.73 (3.5)	0.81 (8)	0.77 (5.5)	0.77 (5.5)	0.67 (1)	0.69 (2)	0.81 (8)	0.77 (5.5)	0.77 (5.5)	0.67 (1)	0.67 (1)	0.69 (2)	0.81 (8)	0.77 (5.5)	0.77 (5.5)	0.67 (1)	0.67 (1)	0.69 (2)	
5	1.03 (11)	0.97 (10)	0.93 (9)	0.84 (6)	0.79 (4)	0.80 (5)	0.85 (7)	0.86 (8)	0.73 (1)	0.78 (3)	0.74 (2)	0.85 (7)	0.86 (8)	0.73 (1)	0.78 (3)	0.78 (3)	0.74 (2)	0.85 (7)	0.86 (8)	0.73 (1)	0.78 (3)	0.78 (3)	0.74 (2)	
6	0.85 (11)	0.78 (9)	0.64 (2)	0.80 (10)	0.72 (6.5)	0.71 (5)	0.72 (6.5)	0.77 (8)	0.62 (1)	0.67 (3)	0.70 (4)	0.72 (6.5)	0.77 (8)	0.62 (1)	0.67 (3)	0.67 (3)	0.70 (4)	0.72 (6.5)	0.77 (8)	0.62 (1)	0.67 (3)	0.67 (3)	0.70 (4)	
7	1.18 (11)	0.92 (8)	0.82 (1)	0.95 (9)	0.87 (6)	0.88 (7)	0.98 (10)	0.86 (4.5)	0.86 (4.5)	0.84 (2.5)	0.84 (2.5)	0.98 (10)	0.86 (4.5)	0.86 (4.5)	0.84 (2.5)	0.84 (2.5)	0.84 (2.5)	0.98 (10)	0.86 (4.5)	0.86 (4.5)	0.84 (2.5)	0.84 (2.5)	0.84 (2.5)	
8	1.16 (11)	1.01 (9)	1.07 (10)	0.95 (8)	0.91 (6)	0.87 (4)	0.91 (6)	0.86 (3)	0.91 (6)	0.81 (2)	0.80 (1)	0.91 (6)	0.86 (3)	0.91 (6)	0.81 (2)	0.81 (2)	0.80 (1)	0.91 (6)	0.86 (3)	0.91 (6)	0.81 (2)	0.81 (2)	0.80 (1)	
9	1.23 (11)	1.17 (9)	1.13 (8)	1.10 (6)	1.05 (5)	1.04 (4)	1.20 (10)	1.12 (7)	0.98 (1.5)	0.99 (3)	0.98 (1.5)	1.20 (10)	1.12 (7)	0.98 (1.5)	0.99 (3)	0.99 (3)	0.98 (1.5)	1.20 (10)	1.12 (7)	0.98 (1.5)	0.99 (3)	0.99 (3)	0.98 (1.5)	
10	1.07 (11)	0.82 (6)	0.83 (8)	0.85 (9)	0.80 (4)	0.82 (6)	0.88 (10)	0.82 (6)	0.76 (3)	0.75 (2)	0.74 (1)	0.88 (10)	0.82 (6)	0.76 (3)	0.75 (2)	0.75 (2)	0.74 (1)	0.88 (10)	0.82 (6)	0.76 (3)	0.75 (2)	0.75 (2)	0.74 (1)	
11	0.95 (11)	0.90 (10)	0.83 (8)	0.82 (7)	0.76 (3.5)	0.76 (3.5)	0.85 (9)	0.77 (5)	0.78 (6)	0.71 (1)	0.73 (2)	0.85 (9)	0.77 (5)	0.78 (6)	0.71 (1)	0.71 (1)	0.73 (2)	0.85 (9)	0.77 (5)	0.78 (6)	0.71 (1)	0.71 (1)	0.73 (2)	
12	1.18 (11)	1.01 (9)	0.92 (6)	0.97 (8)	0.91 (5)	0.90 (4)	1.02 (10)	0.94 (7)	0.88 (1)	0.89 (2.5)	0.89 (2.5)	1.02 (10)	0.94 (7)	0.88 (1)	0.89 (2.5)	0.89 (2.5)	0.89 (2.5)	1.02 (10)	0.94 (7)	0.88 (1)	0.89 (2.5)	0.89 (2.5)	0.89 (2.5)	
13	1.22 (7.5)	1.22 (7.5)	1.24 (9.5)	1.12 (5.5)	1.04 (2)	1.04 (2)	1.24 (9.5)	1.30 (11)	1.12 (5.5)	1.04 (2)	1.07 (4)	1.24 (9.5)	1.30 (11)	1.12 (5.5)	1.04 (2)	1.04 (2)	1.07 (4)	1.24 (9.5)	1.30 (11)	1.12 (5.5)	1.04 (2)	1.04 (2)	1.07 (4)	
14	1.21 (10)	1.31 (11)	1.12 (7.5)	1.08 (5)	1.02 (2)	1.03 (3)	1.12 (7.5)	1.15 (9)	1.10 (6)	1.04 (4)	0.99 (1)	1.12 (7.5)	1.15 (9)	1.10 (6)	1.04 (4)	1.04 (4)	0.99 (1)	1.12 (7.5)	1.15 (9)	1.10 (6)	1.04 (4)	1.04 (4)	0.99 (1)	
15	0.73 (10)	0.74 (11)	0.62 (4)	0.66 (7.5)	0.64 (6)	0.63 (5)	0.67 (9)	0.66 (7.5)	0.57 (3)	0.56 (2)	0.54 (1)	0.67 (9)	0.66 (7.5)	0.57 (3)	0.56 (2)	0.56 (2)	0.54 (1)	0.67 (9)	0.66 (7.5)	0.57 (3)	0.56 (2)	0.56 (2)	0.54 (1)	
16	1.00 (11)	0.91 (10)	0.90 (9)	0.86 (7.5)	0.79 (3.5)	0.79 (3.5)	0.84 (6)	0.86 (7.5)	0.81 (5)	0.74 (2)	0.70 (1)	0.84 (6)	0.86 (7.5)	0.81 (5)	0.74 (2)	0.74 (2)	0.70 (1)	0.84 (6)	0.86 (7.5)	0.81 (5)	0.74 (2)	0.74 (2)	0.70 (1)	
17	0.52 (8)	0.55 (10)	0.55 (10)	0.55 (10)	0.49 (4.5)	0.49 (4.5)	0.50 (6.5)	0.50 (6.5)	0.46 (1)	0.48 (3)	0.47 (2)	0.50 (6.5)	0.50 (6.5)	0.46 (1)	0.48 (3)	0.48 (3)	0.47 (2)	0.50 (6.5)	0.50 (6.5)	0.46 (1)	0.48 (3)	0.48 (3)	0.47 (2)	
A. r.	(10.21)	(9.26)	(7.03)	(7.56)	(4.12)	(4.06)	(8.41)	(7.18)	(3.24)	(2.74)	(2.21)	(8.41)	(7.18)	(3.24)	(2.74)	(2.74)	(2.21)	(8.41)	(7.18)	(3.24)	(2.74)	(2.74)	(2.21)	(2.21)

Table 5: Linear loss for male survey respondents considering either knowledge (know.), attitude (att.) or emotion (emot.) components. The critical differences for the Nemenyi test are: $CD_{p=0.05} = 2.18$ and $CD_{p=0.1} = 1.99$

	Know.	Att.	Emot.	Know./Att	Know./Emot.	Att./Emot.	All
1	1.22 (7)	1.11 (2)	1.12 (3)	1.16 (6)	1.14 (5)	1.08 (1)	1.13 (4)
2	0.86 (4)	0.88 (5)	1.01 (7)	0.83 (3)	0.80 (1)	0.97 (6)	0.82 (2)
3	1.04 (6.5)	1.01 (2)	1.04 (6.5)	1.02 (4.5)	1.02 (4.5)	1.01 (2)	1.01 (2)
4	0.78 (3.5)	0.85 (6)	0.92 (7)	0.78 (3.5)	0.77 (1.5)	0.84 (5)	0.77 (1.5)
5	1.00 (5)	1.03 (7)	1.01 (6)	0.94 (3)	0.92 (2)	0.95 (4)	0.88 (1)
6	0.74 (1)	0.80 (4.5)	0.83 (7)	0.77 (2.5)	0.80 (4.5)	0.82 (6)	0.77 (2.5)
7	1.20 (4)	1.14 (2.5)	1.23 (6)	1.10 (1)	1.25 (7)	1.22 (5)	1.14 (2.5)
8	1.30 (7)	1.14 (6)	1.11 (5)	1.10 (4)	1.05 (2)	1.06 (3)	1.03 (1)
9	1.16 (7)	1.07 (4.5)	1.07 (4.5)	1.10 (6)	1.05 (3)	0.99 (1)	1.00 (2)
10	0.94 (7)	0.75 (1)	0.87 (6)	0.76 (2)	0.84 (5)	0.78 (3)	0.80 (4)
11	0.86 (2.5)	0.87 (4)	1.03 (7)	0.88 (5.5)	0.86 (2.5)	0.88 (5.5)	0.82 (1)
12	0.85 (5)	0.87 (6)	0.91 (7)	0.80 (1)	0.81 (2)	0.82 (3)	0.83 (4)
13	1.10 (5.5)	1.10 (5.5)	1.25 (7)	1.02 (2)	1.06 (4)	1.05 (3)	0.99 (1)
14	1.20 (7)	1.10 (4)	1.13 (6)	1.04 (3)	1.11 (5)	1.02 (2)	1.01 (1)
15	0.92 (6.5)	0.87 (2)	0.91 (5)	0.90 (4)	0.88 (3)	0.86 (1)	0.92 (6.5)
16	0.83 (6.5)	0.68 (2)	0.82 (5)	0.72 (4)	0.83 (6.5)	0.68 (2)	0.68 (2)
17	0.72 (6)	0.69 (3)	0.87 (7)	0.63 (1)	0.70 (4.5)	0.70 (4.5)	0.66 (2)
A. r.	(5.35)	(3.94)	(6.00)	(3.29)	(3.71)	(3.35)	(2.35)

Table 6: Linear loss for female survey respondents considering either knowledge (know.), attitude (att.) or emotion (emot.) components. The critical differences for the Nemenyi test are: $CD_{p=0.05} = 2.18$ and $CD_{p=0.1} = 1.99$

	Know.	Att.	Emot.	Know./Att	Know./Emot.	Att./Emot.	All
1	1.02 (7)	0.88 (1)	0.95 (5)	0.93 (4)	0.97 (6)	0.91 (2)	0.92 (3)
2	0.85 (3)	0.90 (5)	0.91 (6.5)	0.81 (1)	0.87 (4)	0.91 (6.5)	0.84 (2)
3	1.08 (7)	0.89 (3)	0.97 (5)	0.91 (4)	0.99 (6)	0.88 (2)	0.87 (1)
4	0.76 (5)	0.80 (6)	0.88 (7)	0.73 (2)	0.75 (4)	0.74 (3)	0.69 (1)
5	0.93 (6)	0.82 (3)	1.02 (7)	0.75 (2)	0.83 (4)	0.84 (5)	0.74 (1)
6	0.76 (7)	0.67 (2.5)	0.70 (4.5)	0.75 (6)	0.67 (2.5)	0.63 (1)	0.70 (4.5)
7	1.17 (7)	0.85 (4)	1.14 (6)	0.82 (1.5)	1.09 (5)	0.82 (1.5)	0.84 (3)
8	1.13 (7)	0.88 (3.5)	0.95 (6)	0.88 (3.5)	0.91 (5)	0.83 (2)	0.80 (1)
9	1.23 (7)	1.15 (6)	1.14 (5)	0.98 (1.5)	1.04 (4)	1.03 (3)	0.98 (1.5)
10	0.85 (6)	0.82 (4.5)	0.91 (7)	0.81 (3)	0.82 (4.5)	0.79 (2)	0.74 (1)
11	0.89 (7)	0.75 (5)	0.72 (3)	0.71 (2)	0.77 (6)	0.70 (1)	0.73 (4)
12	1.12 (7)	0.91 (2.5)	0.97 (6)	0.96 (5)	0.95 (4)	0.91 (2.5)	0.89 (1)
13	1.23 (6)	1.02 (2)	1.28 (7)	1.00 (1)	1.20 (5)	1.06 (3)	1.07 (4)
14	1.06 (4)	1.08 (5)	1.21 (7)	0.99 (1.5)	1.13 (6)	1.05 (3)	0.99 (1.5)
15	0.73 (7)	0.63 (4.5)	0.66 (6)	0.59 (2.5)	0.63 (4.5)	0.59 (2.5)	0.54 (1)
16	1.02 (7)	0.81 (4)	0.98 (6)	0.77 (2)	0.86 (5)	0.78 (3)	0.70 (1)
17	0.51 (5.5)	0.52 (7)	0.45 (2)	0.46 (3)	0.43 (1)	0.51 (5.5)	0.47 (4)
A. r.	(6.21)	(4.03)	(5.65)	(2.68)	(4.50)	(2.85)	(2.09)