The CSI Journal on
Computer Science and Engineering
Vol. 16, No. 2, 2019
Pages 30-37
Regular Paper

Future Parameter Prediction By Scoring Rules

Saeed Hajizadeh[1], MohammadAmin Fazli[1], Jafar Habibi[1]

[1]Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

Abstract

Scoring rules are common methods for incentivizing experts to present the opinions consistent with their beliefs. Information markets such as prediction markets and decision markets use scoring rules for eliciting the most accurate predictions experts can make. In these markets, experts are invited to buy and sell contracts, according to which they will be paid if their prediction about a future uncertain event is true. The final trading prices can be interpreted as an aggregation of their prediction for an uncertain future event.

In this paper, we propose new mechanisms (prediction oriented and decision oriented) for predicting the value of uncertain continuous variables in the future. These mechanisms in their basic form, are scoring rules with a new paradigm. This paper also includes the results of performing a case study (a prediction oriented mechanism) for predicting the outcome of the 11th presidential election of Iran in 2013 to analyze the performance of our model. Besides the fact that the mechanism's average absolute error in predicting nominees' percentages was low, about 6.53%, it also predicted the final outcome order of all nominees correctly.

Keywords: Scoring Rule, Decision Scoring Rule, Variable Prediction, Election Prediction.

1 Introduction

One of the important aspects of the decision making process in many contexts is to predict the future. Future prediction is a difficult task especially when human participants are involved, as the number of affecting parameters in these environments is very high. A common approach to confront this issue is to extract the relevant information from environment experts and then aggregate this information in order to provide a collective prediction.

In such contexts, a general approach is to invite a set of opinionators in order to capture their predictions about the future and aggregate these opinions into a final prediction. Hanson [1], did a comprehensive study on all information elicitation and aggregation mechanisms which were proposed before, and classified them into two broad categories: "Scoring Rules" and "Information Markets" (or "Prediction Markets"). He also introduced a new category "Market Scoring Rule" that is a combination of scoring rules and information markets. In Section 2, we will explain these mechanisms with details.

Hanson [2] introduced a manipulated version of information markets which he called "Decision Markets" in order to directly predict the consequences of decision makers' taken decisions. Hanson's seminal paper initiated the idea of using prediction mechanisms in decision making where opinions are conditional predictions if different decisions are made.

All studies in this area including different scoring rules, information markets and market scoring rules, are centered around predicting occurrence probabilities for a set of discrete outcomes [1-12]. In this paper new models are introduced to predict the expected value of a continuous variable in future such as net value profit, based on the models and methods used in the literature. The predictions of experts are elicited in templates like "the net profit value will be x", in which x is an amount in the acceptable range of the net profit value. In this paper, both prediction and decision mechanisms are discussed.

One may say that since this prediction task can be interpreted as predicting the occurrence probability of infinite number of outcomes, previous methods can be applied. The difference is, there is an order between outcomes when we are predicting the value of a future continuous variable. For example if someone predicts that the net profit value of the project will be 100, her benefit should differ when the observed value becomes 120 or 200. In our model this property is modeled by the "Validity" feature.

We prove that trivial scoring rules such as negative absolute error can be applied for such predictions as a truthful and valid mechanism. As a case study we ran a game base on these trivial scoring functions to predict the results of Islamic Republic of Iran's 11th presidential election. This game predicted the outcome order of nominees correctly and its average absolute error for nominees' vote shares was low about 6.5%. We also show that trivial scoring functions are not truthful in decision contexts. Finally, we derive some properties of truthful scoring functions which can be used in decision contexts and propose a truthful scoring rule for such contexts .

This paper is organized as follows. In Section 2, "Scoring Rules", "Information Markets" and "Market Scoring Rule" mechanisms are formally described and their related literature is reviewed. In Section 3, our theoretical and experimental results are expressed.

2 Notation & Background

In this section, we will explain mechanisms used for gathering true information from our team of opinionators. These mechanisms are categorized into three main classes "Scoring Rules", "Information Markets" and "Market Scoring Rules". Each of these mechanisms can be used in prediction or decision contexts. Table 1 consists of an overview of the seminal works in these areas.

Table 1: Different methods used for prediction and decision

	Prediction	Decision
Scoring Rules	Square Score Function [13], Logarithmic Score Function [10]	Method Used in [7]
Information Markets	Prediction Markets & Conditional Prediction Markets [4],[14],[15],[16]	Decision Market [2]
Market Scoring Rules	Prediction Market Scoring Rule [1]	Decision Market Scoring Rule [1]

2.1 Scoring Rule

The purpose of scoring rule based mechanisms is to predict the probability of occurrence for a set of outcomes $O = \{o_1, o_2, \ldots, o_n\}$, which are finite, mutually exclusive and covering i.e. these outcomes divide event space into n partitions. More formally,

$$\forall_{i,j\in\{1,2,\ldots,n\},i\neq j}: p(o_i \wedge o_j) = 0$$
$$p(o_1 \vee o_2 \vee \ldots \vee o_n) = 1,$$

When $p(x \wedge y)$ is the probability of occurring both x and y, and $p(x \vee y)$ is the probability of that at least one of them will occur. One of these outcomes will occur and we will name it o^*.

These mechanisms are used in many contexts such as economics [17], weather [18], risk analysis [19], and intelligent computer systems engineering [20] for prediction tasks.

Let $\Delta(O)$ be the set of all possible probability distributions on members of O. For instance $< p_1, p_2, \ldots, p_n >$ can be a member of $\Delta(O)$, if $\sum_i p_i = 1$ and $\forall_i: p_i > 0$. Each participant has a prediction P consistent with her belief which is a vector with n components. The participant believes that the probability of occurrence of o_i is P_i. Assume that she reports her belief with another n-components vector r which is not necessarily equal to P.

Scoring rules are devised to incentivize participants to report truthful predictions. A scoring rule $s: O \times \Delta(O) \rightarrow \mathbb{R}$ is a function of o^* and r which maps them to a real number. So each participant's score can be calculated after observing the future.

The participant's expectation of her score by reporting r is denoted by U(r) which is

$$U(r) = \sum_{i=1}^{n} P_i \times s(o_i, r)$$

Regularity and properness of scoring functions show their correctness and truthfulness [6], [7]. A scoring function is regular if it maps a participant's report to $-\infty$ only if she reports the occurrence probability of observed outcome o_i to zero, that is,

$$s(o_i, r) = -\infty \Rightarrow r_i = 0$$

A proper scoring function won't incentivize a participant to report a prediction inconsistent with her belief. In other words, a scoring function is proper if there is no report which gets more expected score than the participant's true prediction. So if s is proper, we have,

$$\forall_{r\in\Delta(O)} U(P) \geq U(r).$$

A scoring function is strictly proper if it incentivizes participants to report a prediction consistent with their belief. In other words, using a strictly proper scoring rule, the participant's expected score has the maximum value only if the reported value is declared truthfully. s is strictly proper, if the inequality stated in the above condition is strict unless P=r.

Two basic classes of scoring rules are square and logarithmic scores. Square scoring rules are defined in [13]:

$$s(o_i, r) = a_i + b(2r_i - \sum r_i^2)$$

Good [10] introduced logarithmic scoring rules:

$$s(o_i, r) = a_i + b\log(r_i)$$

Binary predictions are prediction tasks in which the set of outcomes O, has only two members such as success and failure. In a scoring rule mechanism for binary prediction tasks, participants need to predict the probability of just one outcome, for example success.

As Winkler [11], [12] defined, binary prediction scoring rule mechanisms consist of two functions $f, g: (0,1) \rightarrow \mathbb{R}$$ for which $f(p)$ and $g(p)$ show the participant's score if success or failure happen respectively.

Logarithmic and square scoring rule can be used in binary predictions. For example in below equations, the first is a binary form of logarithmic scoring rule and the second is a binary form of square scoring rule.

$$f(p) = 1 - (1-p)^2 \ , \qquad g(p) = 1 - p^2$$
$$f(p) = log(p) \ , \qquad g(p) = log(1-p)$$

After eliciting all participants' predictions, their reported values should be aggregated into one final prediction. Averaging is the most common and easiest way, but Hanson [1] expressed its deficiencies and counted the problem of finding a good aggregation method as the flaw of scoring rule mechanisms.

In decision contexts, we have a set of actions $A = \{a_1, a_2, \dots a_m\}$ $ the goal is to predict the probability of different outcomes if each of these actions are taken. So each participant has a belief $P \in \Delta(O)^m$ and expresses a report $r \in \Delta(O)^m$. P_i^j and r_i^j represent occurrence probabilities of o_i with the precondition of choosing a_j, in the participant's belief and report respectively.

After gathering all reports they will be aggregated to one final prediction $B \in \Delta(O)^m$. Using B, we will choose one of the actions to take. "Decision Rule" is a function which shows how to choose an action using B. Decision rules can be divided in two types, deterministic decision rules and stochastic ones. Deterministic decision rules map B to exactly one of the actions, but stochastic ones (which are also called randomized decision rules) determine a probability distribution on the action set, so we choose an action randomly according to that probability distribution.

A deterministic decision rule $D: \Delta(O)^m \rightarrow A$ is a function that maps the aggregated report to one of the possible actions. Deterministic rules are a specific type of stochastic rules.

Othman [7] explained that a decision context prediction mechanism is Independent of Irrelevant Alternatives (IIA) if the expected score depends only on the prediction of the chosen action. More formally in an IIA mechanism, for all $q, r \in \Delta(O)^m$, if $D(q) = D(r) = a_i$ and $q^i = r^i$, we have $U(q) = U(r)$.

Since a truthful mechanism with a deterministic decision rule has to be IIA [7], in such mechanisms the scoring function $S: O \times \Delta(O) \rightarrow \mathbb{R}$ (which is also called "Decision Score Function") maps observed outcome and the reported prediction on the condition of choosing the selected action a_s to a real number.

Binary mechanisms are also studied in the decision context. For example Chen [21] proposed the following pair of strictly proper scoring and stochastic decision rules:

$$f(a_s, r, d) = \sum_j 2r_{success}^j - {r_{success}^j}^2,$$

$$g(a_s, r, d) = -\sum_j {r_{success}^j}^2$$

$$D_i(B) = \frac{B_{success}^i}{\sum_j B_{success}^j}$$

Othman and Sandholm [7] focused on the Max decision rule in binary decision mechanisms. Max chooses the action with the maximum predicted success probability. They showed that symmetric scoring rules can't fulfill truthfulness besides the Max decision rule. Symmetric scoring rules are a pair of functions f and g in which:

$$f(p) = g(1 - p).$$

It is shown that logarithmic and square scoring rules are symmetric [12].

Another category of binary scoring rules is the category of asymmetric rules which are proposed by Winkler [12]. Othman and Sandholm [7] proposed an asymmetric scoring rule which fulfills truthfulness when used with the Max decision rule.

2.2 Information Markets

Prediction markets are quantitative and normative mechanisms to aggregate experts' opinions [22]. In this approach a market is designed in which participants are invited to buy and sell contracts, according to which they will be paid if their prediction about a future uncertain event is true. Finally, the market's trading prices can be interpreted as a collective prediction for that event .

A prediction market can be held with three kinds of contracts "Winner Takes all", "Index" and "Spread" [3] whose details can be found in Table 2.

Table 2: Different types of prediction markets

	Contract Form	What Market Predicts
Winner Takes All	Pays 1$ if o occurs	The final trading prices of contracts predict the occurrence probability of o
Index	Pays x$ if the actual value of variable o becomes x ($0 \leq x \leq 1$)	The final trading prices of contracts predict the mean value of o
Spread	Pays 2$ if $o \geq x$. Contract prices are fixed to 1$.	The amount of x predicts the median value for x.

Markets with "Winner Takes All" contract are studied more than other markets in the literature. The main goal of these markets is to predict the probability distribution of a future event. For example in a football match market participants express three probabilities for the three possible outcomes of the match (win, lose or draw). Prediction markets are proved to be efficient mechanisms for predicting results of elections [4], sport matches [5] and total gross of Hollywood movies [23].

There are a couple of studies on the truthfulness of these markets and their applications [3], [4], [6], [7], [8], [14], [24]. Plott [16] showed that these markets can elicit information distributed between experts and then showed that they can also be used in situations in which we need to elicit and gather information from specific experts. Hanson [15] introduced the use of these markets in politics and elections. He then introduced new ways in which these markets can be useful.

Hanson [2] proposed a new kind of information market which he called "Decision Markets" in order to directly predict the consequences of decision makers' taken decisions. Hanson's idea was to define these markets as a set of conditional prediction markets, one for each possible decision. Othman and Sandholm [7] further represent the difference of a conditional prediction market with a decision market by two words, "Camera" and "Engine". A conditional prediction market is just an effort to predict the occurrence probability of a set of outcomes on condition of the occurrence of one member of a set of events. So it's just like a camera with no effect on the future events, although it tries to predict the future. But a mechanism in decision context is like an engine. The prediction which resulted in this approach will affect which event will occur (which decision to be take) .

Berg and Rietz [24] proved the ability of information markets to be used as decision support systems. They discuss this ability with a famous prediction market which was held 1996 by Iowa Electronic Market (IEM).

2.3 Market Scoring Rule

Hanson [1] introduced a new category "Market Scoring Rule" that is a combination of scoring rules and information markets.

Chen [6] modeled market scoring rules as a sequential game with definite number of turns .

According to his model, each turn one of the participants reports her prediction and changes the aggregated prediction by her reported opinion. After observing the actual outcome she gets the difference she made in the score of aggregated prediction. She will earn up if she enhances the score of aggregated prediction and she will incur losses otherwise.

Hanson [1] reformed this game to a market by proposing "Pricing Rule" which is the inverse of a score function. In this market all participants will buy contracts from market organizers and sell contracts to them. The price of each trade will be specified by pricing rules.

Othman and Sandholm [7] showed that market scoring rule mechanisms in a decision context with deterministic decision rules cannot be truthful. They showed that for any deterministic decision rule and any scoring rule, there can be states that a participant gets higher expected scores by telling a lie, and after these misinforming reports, no one will be incentivized to repair the market's aggregated report.

In some studies like [8], [9], [21] outside cost or benefit was analyzed, such as a payoff from outside the prediction mechanism to practitioners to make them report untruthfully. In this paper, we assume that there is no outside cost or benefit.

3 Results

3.1 Theoretical Results

Assume that X is the target random variable which can take continuous bounded values. X has a probability distribution.

Without loss of generality we can assume that $X \in [0,1]$ which is shown by R^* throughout this paper. Each participant has a notion about the PDF[1] of X and her goal is to predict its outcome of X by reporting its expected value. Let X_p be the notion of the participant about X and $f_{X_p}(t)$ be its probability distribution. Define $r \in R^*$ to be the reported value of this participant. If r is truthful, we should have $r = E\{X_p\}$.

The score of all reports will be calculated after game finished and the real value of X is observed. We define $o \in R^*$ to be observed value of X. Here we have a score function $s: R^* \times R^* \to \mathbb{R}$ for calculating scores from o and r. Each participant has an expectation of her score which is:

$$U(r) = \int_{R^*} f_{X_p}(t) s(t,r) dt$$

We define a new feature, validity, for score functions. Validity refers to the fact that our outcomes are ordered. For instance if one had a closer prediction to the observed value than another's prediction, her report should gain a higher score. A score function is valid if it is strictly descending in $|r-o|$. For a valid scoring function s, we have:

$$\forall_{r_1, r_2 \in R^*}: |r_1 - o| > |r_2 - o| \Rightarrow s(o, r_1) < s(o, r_2)$$

The score function s is proper, if it does not incentivize participants to present a false report on the expected value. It is strictly proper, if it incentivize them to report $E\{X_p\}$. Formally our score function is proper, if for all $r \in R^*$:

$$U(E\{X_p\}) \geq U(r)$$

or

$$\int f_{X_p}(t) s(t, E\{X_p\}) dt \geq \int f_{X_p}(t) s(t,r) dt$$

and if the equality happens only for $r = E\{X_p\}$, the score function is strictly proper.

First, we consider simple observations about the validity and properness of two trivial and common scoring functions "Negative Absolute Error" and "Square". Negative absolute error is defined as follows:

$$s(o,r) = -|r-o|.$$

Observation 1: Negative absolute error is a valid and strictly proper score function.

Proof: We have:

$$|r_1 - o| > |r_2 - o| \Rightarrow -|r_1 - o| < -|r_2 - o|,$$

So, negative absolute error score function is valid.

The expected score for a participant is:

$$U(r) = \int_{R^*} f_{X_p}(t) s(t,r) dt$$
$$= \int_{R^*} f_{X_p}(t)(-|r-o|) dt$$
$$= -(\int_{R^*} f_{X_p}(t)|r-o| dt)$$

Since the result of the final integral is equal or greater than zero, $max_{U(r)}$ is zero. $U(r)$ has its biggest value when $|r - o| = 0$ or $r = o$. □

Square scoring function is defined as:

$$s(o,r) = -(o-r)^2.$$

Observation 2: Square score function is a valid and strictly proper score function.

For proving validity, we have:

$$|r_1 - o| > |r_2 - 0| \Rightarrow (r_1 - o)^2 > (r_2 - o)^2 \Rightarrow -(r_1 - o)^2 < -(r_2 - o)^2.$$

So it is valid. To discuss being strictly proper, we should calculate $U(r)$:

$$U(r) = \int f_{X_p}(t)(-(t-r)^2) dt$$
$$= -(\int \left\{t^2 f_{X_p}(t) dt\right\} + r\text{^}2 - 2rE\{X_p\}).$$

Thus, we have:

$$\frac{d}{dr} U(r) = -2r - rE\{X_p\} = 0 \Rightarrow r = E\{X_p\}.$$

Since $U(r)$ is concave, the participant will have her higher expected score only when she reports the truth.

We will show the performance of trivial scoring functions in predicting a continuous future variable by running a case study (See Subsection 3.2 for more details). Now let's switch to decision contexts. In decision contexts, proper and strictly proper refers to a pair of a score function and a decision rule. Consider the set of possible actions $A = \{a_1, a_2, \ldots, a_m\}$. We represent the variable X in the case of choosing a_i by X^i. Each participant has a belief about X which is shown by X_P. X_P^i means the belief of the participant about X in case of choosing a_i . In other word each participant estimates m PDFs like $f_{X_P^i}(t)$. They will report $r \in R^{*m}$ and it is truthful when:

$$\forall_{1 \leq i \leq m}\, r^i = E\{X_P^i\}$$

After information gathering finishes, we will aggregate participants' reports to a final report $b \in R^{*m}$. Decision rules

[1] Probability Distribution Function

will choose one of actions from b. Decision rules can be deterministic ($D: R^{*m} \rightarrow A$) or stochastic ($D: R^{*m} \rightarrow \Delta(A)$) . Let a_s be the selected action. The score of a report will be calculated after observing the actual value of X which we show by $o \in R^*$. So the scoring function $S: R^* \times R^* \rightarrow \mathbb{R}$$ is a function of o and r^s. The participant expected score can be computed as the following:

$$U(r) = \int_{R^*} f_{X_P^{D(r)}}(t)\, S(t, r^{D(r)})\, dt.$$

Othman and Sandholm [7] proved that in predicting the occurrence probabilities of a set of discrete outcomes, there is no strictly proper scoring rules for a mechanism with a deterministic decision rule. Theorem 1 proves a similar result for mechanisms for predicting continuous variables.

Theorem 1: There is no strictly proper scoring rule for a mechanism for predicting continuous variables with a deterministic decision rule.

Proof: Suppose $r_1, r_2 \in R^*$ are two different reports such that:

$$D(r_1) = D(r_2) = a_s, r_1^s = r_2^s$$

The expected score for these reports are

$$U(r_1) = \int_{R^*} f_{X_P^s}(t)\, S(t, r_1^s)\, dt$$

and

$$U(r_2) = \int_{R^*} f_{X_P^s}(t)\, S(t, r_2^s)\, dt$$

One can see that since $r_1^s = r_2^s$, for each $t \in R^*$, we have $S(t, r_1^s) = S(t, r_2^s)$. Thus $U(r_1) = U(r_2)$, so we may have two different reports (and one of them is necessarily a false one) that have equal expected score. □

Othman and Sandholm defined another feature named quasi-strictly proper. <D, S> is quasi-strictly proper if the most expected score is for the case that a participant has reported truthfully for the action which will be selected based on her report, and she doesn't misreport in a way that the selected action changes. In our formalism, if <D, S> is quasi-strictly proper, U is the expected score based on S and $r^* \in R^*$ is a truthful report, we have:

$$\forall_{r \in R^*}\, U(r^*) \geq U(r)$$

And the equality happens only when

$$D(r) = D(r^*) = a_s$$

And

$$r^s = r^{*s}$$

Our next result (Theorem 2) is about deriving sufficient conditions for the possibility of designing quasi-strictly mechanisms with one of the most important deterministic decision rules which is called the max rule. Max rule (D_{MAX}) can be defined formally as follows:

$$d_{max}(b) = argmax_i\, b^i$$
$$D_{MAX}(b) = a_{d_{max}(b)}$$

This rule is a common one. As an example consider a case when we choose an action with maximum benefit.

Theorem 2: Following conditions are sufficient for an IIA mechanism <D_{MAX}, S> to be quasi-strictly proper.

1. U(r) can be written as $f(r^{d_{max}(r)})$ where f is a single variable increasing function.
2. U(r) has its maximum value when $r^{d_{max}(r)} = E\{X_P^{d_{max}(r)}\}$ ($r^{d_{max}(r)}$ is truthfully reported)

Proof: Suppose p is a truthful report, r is the stated report and $s = d_{max}(r)$. Assume that <D_{MAX}, S> fulfills the two mentioned conditions. By the definition of quasi-strictly proper pairs, <S, D> is not quasi-strictly proper if and only if one of these situations is met:

1. $r^s \neq p^s$ and $U(r) \geq U(p)$
2. $r^s = p^s$ but reporting p results in choosing $a_{s'} \neq a_s$ and $U(r) \geq U(p)$

From truthfulness of p, we know that $p^s = E\{X_P^s\}$. U(r) has its maximum value when $r^s = E\{X_P^s\}$, thus we must have $r^s = p^s$. Therefore situation 1 will never be met.

Now assume that $r^s = p^s$ and reporting p results in choosing $a_{s'}$. Thus $d_{max}(p) = s'$ and $p^{s'} > p^s$. Thus we have:

$$U(r) = f(r^s) = f(p^s).$$

Since f is increasing,

$$U(p) = f(p^{s'}) > f(p^s) = f(r^s) = U(r)$$$.$$

Thus U(p) > U(r) and situation 2 will never be met. □

After proving the last theorem, we introduce a valid scoring function which makes a quasi-strictly proper pair with max decision rule. We call it "Polynomial" scoring function.

$$S_{Poly}(o, r^s) = r^s\,(2o - r^s)$$

Observation3: S_{Poly} is valid and <D_{MAX}, S_{Poly}> is a quasi-strictly proper mechanism.

Proof: Since validity of polynomial score function is straightforward, we just need to prove that <D_{MAX}, S_{Poly}> is quasi-strictly proper. For this we should prove that S_{Poly} meets the two conditions considered in Theorem 2. Assuming that $s = d_{\max}(r)$, we have:

$$U(r) = \int_{R^*} f_{X_P^s}(t) S_{Poly}(t, r^s) dt$$
$$= 2r^s \int_{R^*} t\, f_{X_P^s}(t) dt \; - r^{s^2} \int_{R^*} f_{X_P^s}(t) dt$$
$$= 2r^s\, E\{X_P^s\} - r^{s^2}$$

To prove condition 2, it should be proved that the most expected score will be reached when r^s is reported truthfully.

$$\frac{\partial\, U(r)}{\partial\, r^s} = 2E\{X_P^s\} - 2r^s = 0 \Rightarrow r^s = E\{X_P^s\}.$$

To prove the first condition, it is enough to replace $E\{X_P^s\}$ with r^s in

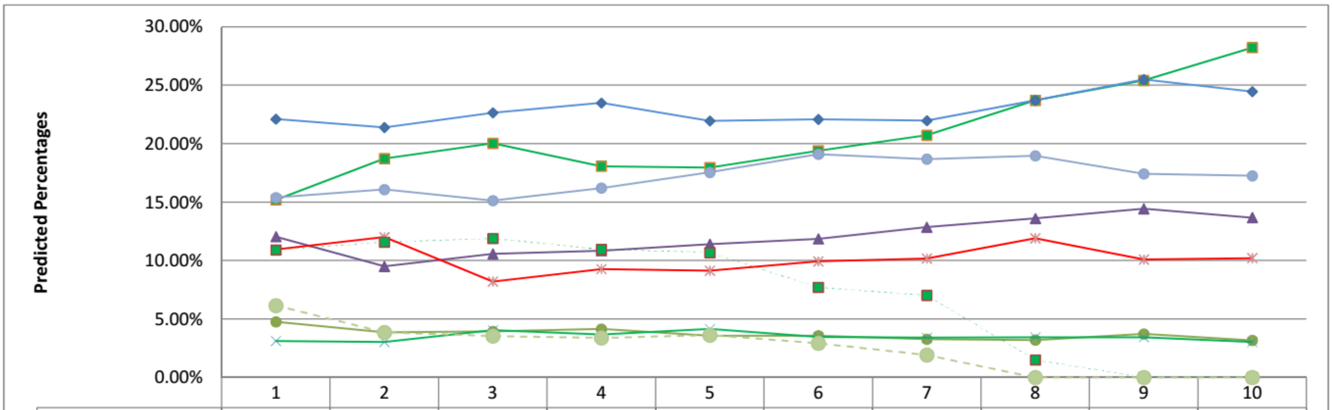
$$U(r) = 2r^s\, E\{X_P^s\} - r^{s^2},$$

We will have $U(r) = r^{s^2}$ which is a strictly increasing function. □

3.2 Case Study

Based on the absolute error score function, a scoring rule mechanism was designed to predict the result of 11th Iranian Presidential Election. This mechanism was tested in an online web site. The project tended to predict election results, using expert predictions. The mechanism elicited participants' opinions and aggregated them with a simple average aggregation.

Based on "Presidential Election Law" of Iran, after exclaiming accredited nominees by the guardian council the election competition starts. The ministry of interior is in charge of holding election and declares the results. In 11th Iranian Presidential Election 8 people were exclaimed as accredited nominees. Mr. Gholamali Haddad-E-Adel and Mr. Mohammad Reza Aref withdrew from running 3 days and 2



	1	2	3	4	5	6	7	8	9	10
Rohani	15.19%	18.73%	20.03%	18.06%	17.96%	19.38%	20.72%	23.69%	25.41%	28.21%
Ghalibaf	22.10%	21.38%	22.65%	23.48%	21.95%	22.07%	21.96%	23.69%	25.49%	24.46%
Jalili	15.38%	16.08%	15.13%	16.20%	17.54%	19.09%	18.68%	18.96%	17.42%	17.25%
Rezai	12.05%	9.50%	10.57%	10.85%	11.40%	11.86%	12.86%	13.60%	14.43%	13.67%
Velayati	10.95%	12.00%	8.21%	9.28%	9.14%	9.93%	10.17%	11.91%	10.09%	10.21%
Gharzi	4.76%	3.85%	3.94%	4.14%	3.56%	3.57%	3.27%	3.20%	3.72%	3.17%
Spoilt Votes	3.12%	3.04%	4.04%	3.66%	4.15%	3.45%	3.40%	3.44%	3.44%	3.03%
Aref	10.90%	11.58%	11.90%	10.95%	10.68%	7.71%	7.03%	1.51%	0	0
Haddad E Adel	6.14%	3.85%	3.55%	3.38%	3.62%	2.93%	1.92%	0	0	0

Figure 1: Prediction trends in 10 games

days respectively before the election, so the final set of nominees in the election had 6 candidates.

Our contest was held in 10 games - each day a game, in a 10 days interval followed by the election. In each game, each participant could expose his/her prediction about the vote share of nominees and spoilt votes. Literally they predicted the percentages which would be exclaimed by Ministry of Interior of Islamic Republic of Iran after the election .

It was declared that after exclaiming the results, all prediction would be scored using negative absolute error score function. Although Schreiber [23] showed that virtual scores can play the role of real money in these mechanisms without any remarkable decrease in accuracy, we declared that the most accurate prediction will be paid a constant payoff.

The project was held between 4/6/2013 and 13/6/2013. Average number of participants in a game was 66. Number of participants of each game is shown in Table 3. Figure 1 shows the prediction trend on 10 games, which is also the prediction changes over a 10-days interval before the election.

Table 3: Number of participants in each game

Game Number	1	2	3	4	5	6	7	8	9	10
Number of Participants	21	26	77	65	78	86	78	75	79	76

The last predictions which was declared on 13/6/2013 at 23:59 is assumed as the final prediction of this project. The final prediction, actual exclaimed percentages and their absolute differences are showed in Table 4.

Table 4: Final Predictions resulted from this project, actual values and absolute errors of predictions

Prediction Variable	Predicted Value	Actual Value	Absolute Error
Vote Percentage of Mr. Rouhani	28.2	50.7	22.5
Vote Percentage of Mr. Ghalibaf	24.5	16.5	7.9
Vote Percentage of Mr. Jalili	17.2	11.4	5.9
Vote Percentage of Mr. Rezaei	13.7	10.6	3.1
Vote Percentage of Mr. Velayati	10.2	6.2	4
Vote Percentage of Mr. Gharazi	3.2	1.2	1.9
Vote Percentage of Spoilt Votes	3	3.4	0.4

Eventually the project predicted the value of these 7 variables with average absolute error of 6.35%.

The predictions became more accurate as we got closer to the Election Day. The trend of prediction errors is depicted in Figure 2.

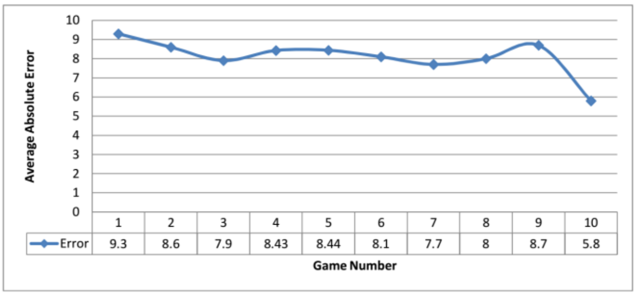


Figure 2: Average absolute error in each game

Comparing Results with Other Works

All other predictions that were taken were based on election polls. A comparative evaluation for this project and other predictions on the presidential election results is done with two error functions: average absolute error and sum of ranking absolute errors (abbreviated with "Ranking Error"). Ranking error shows the sum of ranking absolute errors for each of 6 nominees. Suppose that π^{Actual} is a permutation which shows the actual ranking of nominees i.e. π_i^{Actual} is the rank of the

i[th] nominee. Similarly $\pi^{Preddicted}$ shows the predicted prediction. Ranking error is formally defined as:
$RankingError(\pi^{Actual}, \pi^{Predicted}) = \sum_i |\pi_i^{Actual} - \pi_i^{Predicted}|$.
$Ranking\ Error = 0$ shows that the ranking of all nominees are predicted correctly. Table 5 shows studied polls and their errors.

Table 5: Presidential election polls' details

Poll	Poll Type	Holding Data	# of Participants	Average Absolute Error	Ranking Error
IPOS	Phone	13/6/2013	1067	4.8	2
INN	Field	11/6/2013	12500	3.7	4
Entekhab E Shoma	Online	13/6/2013	15023	7.3	2
Nazar E Akhar	Online	14/6/2013	30989	8.4	2
Asr Iran	Online	12/6/2013	313345	7.6	2
Entekhab	Online	14/6/2013	57020	8.9	0
Yek Ray	Online	14/6/2013	59480	10.7	8
Entekhab at 92	Online	13/6/2013	93418	8.5	3
Khabarpu	Online	14/6/2013	170139	2.5	4

Since some of the predictions just predicted the vote share of nominees (not spoiled votes) we only focus on the variables showing each nominees' vote share. Figures 3 and 4 show the rating of all predictions on the election including this project, which is demonstrated by "Scoring Rule". The ratings are based on average absolute error and ranking error.

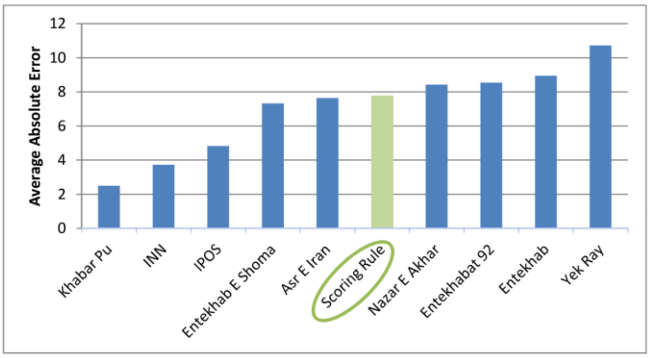


Figure 3: Comparing our method with other predictions based on average absolute error

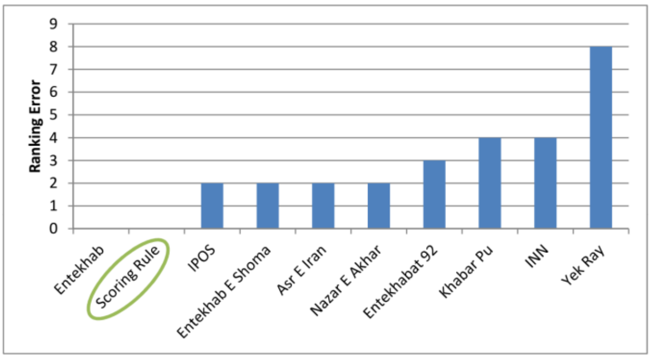


Figure 4: Comparing our method with other predictions based on ranking error

4 Conclusion

In this paper, a study on prediction mechanisms based on the Hanson category of information gathering methods was presented and a new model for predicting the value of a variable and deciding based on that was proposed. This model was defined in a simple prediction context and a decision one. "Validity" as a new feature was defined for score functions. Applicability and truthfulness of trivial scoring rules "Negative Absolute Error" and "Square" in a simple prediction context were proved. It was shown that these trivial scoring rules cannot fulfill truthfulness in decision contexts, and the "Polynomial" score function was defined as an applicable and truthful score function which makes a quasi-strictly pair with "Max" decision rule.

Also a case study was performed to predict Iran's 11th presidential election results. This case study was a scoring rule mechanism using "Negative Absolute Error" score function and average aggregation rule. This project resulted in a prediction with average absolute error of 6.53% in predicting vote percentage of nominees and spoilt votes. It also predicted the ranking of nominees correctly.

References

[1] Hanson. R, "Combinatorial information market design," Information Systems Frontiers, vol. 5, no. 1, pp.107-119.

[2] Hanson. R, Decision markets. Entrepreneurial Economics: Bright Ideas from the Dismal Science, vol. 79, 2002.

[3] Wolfers. J, and Zitzewitz. E, Prediction markets. Journal of economic perspectives, vol. 18, no. 2, pp.107-126, 2004.

[4] Berg. J.E, and Rietz. T.A, Prediction markets as decision support systems. Information systems frontiers, vol. 5, no. 1, pp. 79-93, 2003.

[5] Chen. Y, Chu. C.H, Mullen. T, and Pennock. D.M, "Information markets vs. opinion pools: An empirical comparison," In Proceedings of the 6th ACM conference on Electronic commerce, pp. 58-67, 2005.

[6] Chen. Y, Kash. I, Ruberry. M, and Shnayder. V, "Decision markets with good incentives," In International Workshop on Internet and Network Economics, pp. 72-83, 2011.

[7] Othman. A, and Sandholm. T, "Decision rules and decision markets," In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 625-632, 2010.

[8] Dimitrov. S, and Sami. R, "Composition of markets with conflicting incentives," In Proceedings of the 11th ACM conference on Electronic commerce, pp. 53-62, 2010.

[9] Chen. Y, Gao. X.A, Goldstein. R, and Kash. I.A, "Market manipulation with outside incentives," In Autonomous Agents and Multi-Agent Systems, vol. 29, no. 2, pp.230-265, 2015.

[10] Good. I.J, Rational decisions, Journal of the Royal Statistical Society. Series B (Methodological), pp.107-114, 1952.

[11] Winkler. R.L, "Scoring rules and the evaluation of probability assessors," Journal of the American Statistical Association, vol. 64, no. 327, pp.1073-1078, 1969.

[12] Winkler. R.L, "Evaluating probabilities: Asymmetric scoring rules," Management Science, vol. 40, no. 11, pp. 1395-1405, 1994.

[13] Brier. G.W, "Verification of forecasts expressed in terms of probability," Monthey Weather Review, vol. 78, no. 1, pp. 1-3, 1950.

[14] Ortner. G, Forecasting markets, an industrial application. Mimeo, 1998.
[15] Hanson. R, “Shall we vote on values, but bet on beliefs?” Journal of Political Philosophy, 2003.
[16] Plott. C.R, “Markets as information gathering tools,” Southern Economic Journal, vol. 67, no. 1, pp. 2-15, 2000.
[17] O'Carroll. F.M, “Subjective probabilities and short-term economic forecasts: An empirical investigation,” Applied Statistics, pp. 269-278, 1977.
[18] Murphy. A.H, and Winkler. R.L, “Probability forecasting in meteorology,” Journal of the American Statistical Association, vol. 79, no. 387, pp. 489-500, 1984.
[19] Dewispelare. A.R, Herren. L.T, and Clemen. R.T, “The use of probability elicitation in the high-level nuclear waste regulation program,” International Journal of Forecasting, vol. 11, no. 1, pp.5-24, 1995.
[20] Druzdzel. M.J, and Van Der Gaag. L.C, “Elicitation of probabilities for belief networks: Combining qualitative and quantitative information,” In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pp. 141-148, 1995.
[21] Chen. Y, and Kash. I.A, “Information elicitation for decision making,” In The 10th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 175-182, 2011.
[22] Glenn. J.C, and Gordon. T.J, Futures research methodology: version 2.0. Amer Coin for the Un Univ, 1999.
[23] Servan-Schreiber. E, Wolfers. J, Pennock. D.M, and Galebach. B, “Prediction markets: Does money matter?” Electronic markets, vol. 14, no. 3, pp. 243-251, 2004.
[24] Berg. J, Forsythe. R, Nelson. F, and Rietz. T, “Results from a dozen years of election futures markets research,” Handbook of experimental economics results, vol. 1, pp. 742-751, 2008.

Saeed Hajizadeh is currently the CTO of Ad-Venture Social Network Advertising Agency. He received his BSc and MSc in IT engineering from Tehran University (2011) and Sharif University of Technology (2013) respectively. His research interests concern Social Networks, Prediction and Game Theory.

Email: hajizadeh@ce.sharif.edu

MohammadAmin Fazli received his BSc in hardware engineering and MSc and PhD in software engineering from Sharif University of Technology, in 2009, 2011 and 2015 respectively. He is currently a Lecturer at Sharif University of Technology and R&D Supervisor at Sharif's Intelligent Information Center resided in this university. His research interests include Game Theory, Combinatorial Optimization, Computational Business and Economics, Graphs and Combinatorics, Complex networks and Dynamical Systems.

Email: fazli@sharif.edu

Jafar Habibi is an Associate Professor in the Department of Computer Engineering, Sharif University of Technology, and the managing director of Intelligent Information Solutions Center. He is a supervisor of Sharif’s Robo-Cup Simulation Group and Software Engineering Lab. His research interests are mainly in the areas of software engineering, simulation systems, MIS, DSS, and evaluation of computer systems performance.

Email: jhabibi@sharif.edu

Paper Handling Data:

Submitted: 02.08.2019
Received in revised form: 05.15.2019
Accepted: 06.01.2019
Corresponding author: MohammadAmin Fazli
Computer Engineering Department, Sharif University of Technology, Tehran, Iran