Simulation, learning, and optimization techniques in Watson's game strategies

G. Tesauro
D. C. Gondek
J. Lenchner
J. Fan
J. M. Prager

The game of Jeopardy!™ features four types of strategic decision-making: 1) Daily Double wagering; 2) Final Jeopardy! wagering; 3) selecting the next square when in control of the board; and 4) deciding whether to attempt to answer, i.e., ''buzz in.'' Strategies that properly account for the game state and future event probabilities can yield a huge boost in overall winning chances, when compared with simple ''rule-of-thumb'' strategies. In this paper, we present an approach to developing and testing components to make said strategy decisions, founded upon development of reasonably faithful simulation models of the players and the Jeopardy! game environment. We describe machine learning and Monte Carlo methods used in simulations to optimize the respective strategy algorithms. Application of these methods yielded superhuman game strategies for IBM Watson™ that significantly enhanced its overall competitive record.

Introduction

Major advances in question-answering (QA) technology were needed for IBM Watson* to play Jeopardy!** at a championship level. The game show requires rapid-fire answers to challenging natural-language questions, broad general knowledge, high precision, and accurate confidence estimates. Building a computing system that can answer Jeopardy! questions at a performance level close to that of a human required intense work over a four-year period by a team of two dozen IBM Researchers; this was the primary focus and main technical achievement of the Watson project.

Having pushed QA to a competitive level with humans in the Jeopardy! domain, it became important to focus on the game's strategic aspects in order to maximize Watson's chance of winning. There are four types of strategy decisions [1]: 1) wagering on a Daily Double (DD); 2) wagering during Final Jeopardy! (FJ); 3) selecting the next square when in control of the board; and 4) deciding whether to attempt to answer, i.e., "buzz in." The most critical junctures of a game often occur in DDs and the FJ rounds, where wagering is required. Selecting a judicious amount to wager, on the basis of one's confidence and the specific game situation, can make a substantial difference in a player's overall chance to win. In addition, given the importance of DDs, it follows that a player's square selection strategy should result in a high likelihood of finding a DD. Allowing one's opponents to find the DDs can lead to devastating consequences, particularly when playing against Grand Champions of the caliber of Ken Jennings and Brad Rutter. Furthermore, a contestant's optimal buzz-in strategy can change dramatically in certain endgame scenarios. For example, a player whose score is just below half the leader's score may need to make a "desperation buzz" on the last question to try to avoid a sure loss. Conversely, at just above half the leader's score, the correct strategy may be to never buzz in.

This paper describes our team's work in developing a collection of game-strategy algorithms deployed in Watson's live Jeopardy! contests against human contestants. To our knowledge, this paper presents the first-ever quantitative and comprehensive approach to the Jeopardy! strategy that is explicitly based on estimating and optimizing a player's probability of winning in any given Jeopardy! game state. Our methods enable Watson to find DDs faster than humans and to calculate optimal wagers and buzz-in thresholds to a degree of precision going well beyond human capabilities in live gameplay. Watson's use of these advanced quantitative game strategies significantly enhanced its overall competitive record, as detailed in the following.

Digital Object Identifier: 10.1147/JRD.2012.2188931

0018-8646/12/$5.00 © 2012 IBM

The following section first describes our development of a Jeopardy! simulator, which we use to simulate contests between Watson and human contestants. Building the simulator entailed mining historical data on contestant performance in thousands of previous episodes, to obtain models of the statistical performance profiles of human contestants and their tendencies in wagering and square selection.

We then present four specific methods for designing, learning, and optimizing Watson's four strategy modules over the course of many simulated games. These methods include the following: 1) DD wagering using a combination of nonlinear regression with reinforcement learning; 2) FJ wagering based on a Best Response calculation using extensive Monte Carlo sampling; 3) square selection based on live Bayesian inference calculation of the DD location probabilities; 4) buzz-in thresholds in endgames using a combination of Approximate Dynamic Programming with online Monte Carlo trials.

Jeopardy! simulation model

Since we optimize Watson's strategies over millions of synthetic matches, it is important that the simulations be faithful enough to predict outcome statistics of live matches. Developing such a simulator required significant effort, particularly in the development of human opponent models, which is also critically important in the game of poker [2]. Our simulator uses stochastic models of the various events that can occur at each step of the game, ignoring the language content of category titles, questions, and answers. These models are informed by the following.

a) Properties of the game environment (rules of play, DD placement probabilities, etc.).
b) Performance profiles of human contestants, including tendencies in wagering and square selection.
c) Performance profiles of Watson, along with Watson's actual strategy algorithms.
d) Estimates of relative "buzzability" of Watson versus humans, i.e., how often a player is able to buzz in first or "win the buzz," when all contestants are attempting to buzz in.

Our primary source of information regarding a) and b) is a collection of comprehensive historical game data available on the J! Archive website [3]. We obtained fine-grained event data from approximately 3,000 past episodes, going back to the mid-1990s, annotating the order in which questions were played, right and wrong contestant answers, DD and FJ wagers, and the DD locations.

We devised three different human models, corresponding to different levels of opposition in Watson's matches with human contestants. The Average Contestant model was fitted to all game data, except Tournament of Champions data; this was an appropriate model of Watson's opponents in its first series of sparring games against former Jeopardy! contestants, which took place in late 2009 and early 2010. The Champion model was designed to represent much stronger opponents that Watson faced in a second series of sparring games during Fall 2010: Those contestants had competed in the Tournament of Champions and had reached the final or semifinal rounds. We developed this model using data from the J! Archive list of the 100 best players, ranked according to number of games won. Finally, for our exhibition match with Ken Jennings and Brad Rutter, we devised a Grand Champion model that was informed by performance metrics of the ten best players.

Since the exhibition match used a multigame format (first, second, and third place determined by two-game point totals), we developed specialized DD and FJ wagering models for Games 1 and 2 of the match, as described at the end of this section.

Daily Double placement

Our computations of joint row–column frequencies based on the J! Archive data of Round 1 and Round 2 DD placement confirmed the well-known observations that DDs tend to be found in the lower rows (third, fourth, and fifth) of the board and basically never appear in the top row.

However, we were surprised to discover that some columns are more likely to contain a DD than others. For example, DDs are most likely to appear in the first column and are least likely to appear in the second column. Additional analytic insights from the data include the following: 1) The two second-round DDs never appear in the same column; 2) the row location appears to be set independently of the column location and independently of the rows of other DDs within a game; and 3) the Round 2 column-pair statistics are mostly consistent with independent placement, apart from the constraint in 1). However, there are a few specific column pair frequencies that exhibit borderline statistically significant differences from an independent placement model.

Based on our analysis, the simulator assigns the DD location in Round 1, and the first DD location in Round 2, according to the respective row–column frequencies. The remaining Round 2 DD is assigned a row unconditionally, but its column is assigned conditioned on the first DD column.

Daily Double accuracy/betting model

We set DD accuracy in the Average Contestant model to 64%, from the mean DD accuracy of all contestants over the J! Archive regular episode data set. The DD accuracy in the Champion model rises to 75%, and that in the Grand Champion model rises to 80.5%.

Bets made by human contestants tend to be round-number bets such as $1,000 or $2,000 and rarely exceed $5,000. The main dependencies that we observed are that players in

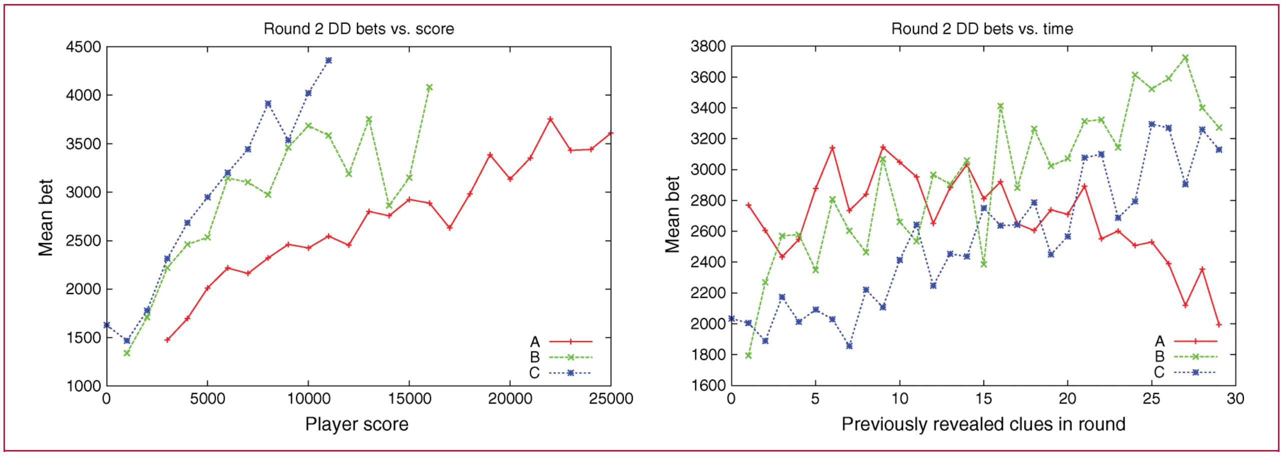

Figure 1

Average Round 2 DD bets of human contestants in (A) first place, (B) second place, and (C) third place. Left: Bets as a function of player score. Right: Bets as a function of clues played in round.

the lead tend to bet more conservatively and become extremely conservative near the end of the game, presumably to protect their lead going into FJ. These dependencies are clearly shown in **Figure 1**, where we plot average bets as functions of player score and of clues played in the second round.

While these observed wagering strategies were built into our Average Contestant model, we surmised (correctly as it turned out) that much stronger Champion and Grand Champion players would quickly realize that they need to bet DDs extremely aggressively when playing against Watson. Thus, these models employed an aggressive heuristic strategy that would bet nearly everything, unless a heuristic formula indicated that the player was close to a mathematically certain win.

Final Jeopardy! accuracy/betting model

Our J! Archive data indicates that average contestants have approximately 50% chance to answer FJ correctly, whereas Champions and Grand Champions, respectively, have approximately 60% and 66% FJ accuracy. Furthermore, from statistics on the eight possible right/wrong triples, it is also clear that right/wrong answers are positively correlated among contestants, with a correlation coefficient $\rho_F \sim 0.3$ providing the best fit to the data. The simulator implements draws of correlated random binary right/wrong outcomes by first generating correlated real numbers using a multivariate normal distribution and then by applying suitably chosen thresholds to convert to 0 or 1 outcomes at the desired mean rates [4].

The most important factor in FJ wagering is score positioning, i.e., whether a player is in first place ("A"), second place ("B"), or third place ("C"). To develop stochastic-process models of likely contestant bets, we first discarded data from "lockout" games (where the leader has a guaranteed win), and then examined numerous scatter plots such as those shown in **Figure 2**. We see a high-density line in A's bets corresponding to the well-known strategy of betting to cover in the case that B's score doubles to 2B. Likewise, there are two high-density lines in the plot of B's bets, one where B bets everything and one where B bets just enough to overtake A. Yet, there is considerable apparent randomization apart from any known deterministic wagering principles.

After a thorough examination, we decided to segment the wagering data for A and B into six groups: We used a three-way split based on strategic breakpoints in B's score relative to A's score (less than two-thirds, between two-thirds and three-fourths, and more than three-fourths), plus a binary split based on whether B has at least double C's score. We then devised wagering models for A, B, and C[1] that choose among various types of betting logic, with probabilities based on observed frequencies in the data groups. As an example, our model for B in the case ($B \geq \frac{3}{4}A$, $B \geq 2C$) bets as follows: bet "bankroll" (i.e., nearly everything) with 26% probability, "keepout C" (i.e., just below B-2C) with 27% probability, "overtake A" (i.e., slightly above A-B) with 15% probability, "two-thirds limit" (i.e., just below 3B-2A) with 8% probability, and various types of random bets with the remaining 24% probability mass.

The betting models described in the following were designed solely to match human bet distributions, and were

[1] Curiously enough, we saw no evidence that C's wagers vary with strategic situation; therefore, we implemented a single betting model for C covering all six groups.

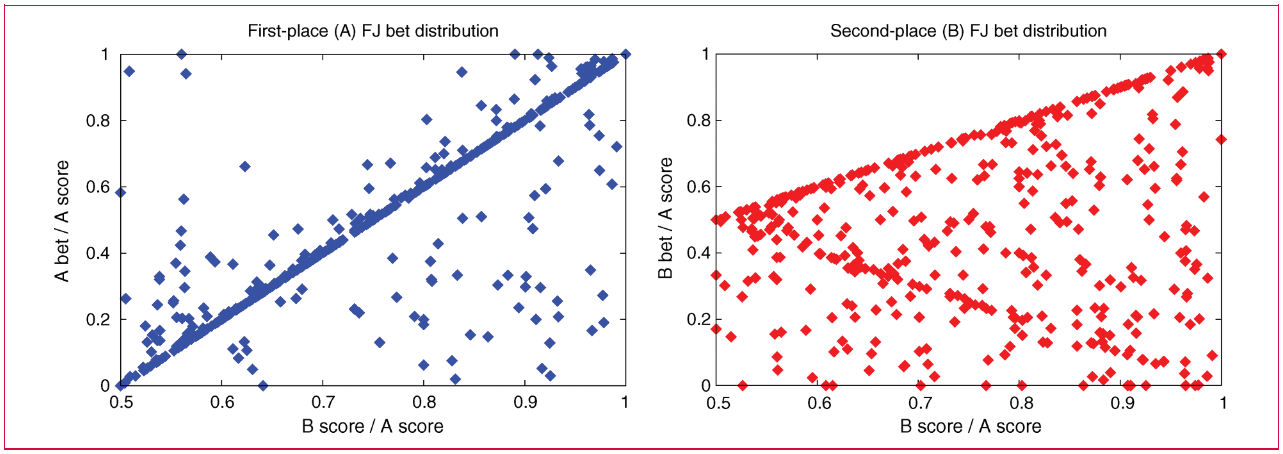

Figure 2

Distribution of human Final Jeopardy! bets, normalized by leader score, as a function of the ratio of second-place/first-place scores. Left: bets of first-place player "A." Right: bets of second-place player "B."

Table 1 Comparison of actual human win rates with model win rates by historic replacement in 2,092 nonlocked Final Jeopardy! situations from past episodes. (A: first place; B: second place; C: third place.)

Role	*Real*	*Model*
A	65.3%	64.8%
B	28.2%	28.1%
C	7.5%	7.4%

not informed by human FJ win rates. However, we subsequently verified by a historical replacement technique that the models track actual human win rates quite closely, as shown in **Table 1**. We first measured the empirical win rates of the A, B, and C roles in 2,092 nonlocked FJ situations from past episodes. We then took turns recalculating the win rate of one role after replacing the bets of that role by the bet distribution of the corresponding model. The models match the target win rates very well, considering that the human bets are likely to reflect unobservable confidence estimates given the FJ category.

Regular question model

Our stochastic-process model of regular questions generates a random correlated binary triple indicating which players attempt to buzz in, and a random correlated binary triple indicating whether the players have a correct answer. In the case of a contested buzz, a buzz winner is randomly selected based on the contestants' relative buzzability (assumed equal in all-human matches). The model, thus, has four tunable parameters: mean buzz attempt rate b, buzz correlation ρ_b, mean precision p, and right/wrong correlation ρ_p.

We set the four parameter values by fitting to observed frequencies of the seven possible outcomes for regular questions, as depicted in **Figure 3**. The outcome statistics are derived from J! Archive records of more than 150,000 regular questions. The resulting parameter values were: $b = 0.61$, $\rho_b = 0.2$, $p = 0.87$, and $\rho_p = 0.2$. The right/wrong correlation is noteworthy: although a positive value is reasonable, given the correlations seen in FJ accuracy, it might be surprising due to the "tip-off" effect on rebounds. When the first player to buzz gives a wrong answer, this eliminates a plausible candidate and could significantly help the rebound buzzer to deduce the correct answer. We surmise that the data may reflect a knowledge correlation of $\sim$0.3 combined with a tip-off effect of ~ -0.1 to produce a net positive correlation of 0.2.

In the Champion model, there is a substantial increase in attempt rate $(b = 0.8)$ and a slight increase in precision $(p = 0.89)$. In the Grand Champion model, we estimated further increases in these values, to $b = 0.855$ and $p = 0.915$, respectively. We also developed refined models where b and p values depended on round and on dollar value; these refinements make the simulations more accurate but do not have a meaningful impact on the optimization of Watson's strategies.

Square selection model

Most human contestants tend to select in top-to-bottom order within a given category and to stay within a category rather than jumping across categories. There is a further weak tendency to select categories moving left-to-right across the board. On the basis of these observations and on the likely

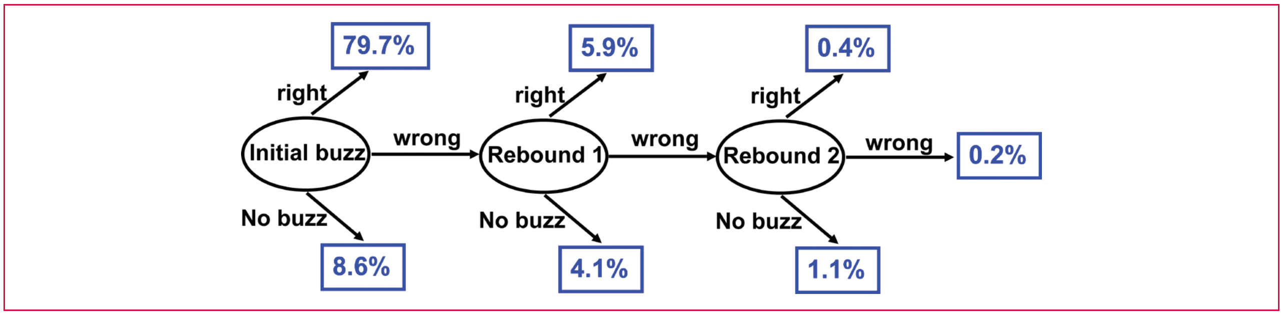

Figure 3

Frequencies in J! Archive data of seven possible outcomes of regular questions.

impact of Watson's square selection, we developed an Average Contestant model of square selection that stays in the current category with 60% probability and otherwise jumps to a random different category. When choosing within a category, there is high probability (~90%) of choosing the topmost available question. By contrast, we model Champion and Grand Champion square selection as DD seeking based on the known row statistics of DD placement. Strong players generally exhibit more DD seeking when selecting squares, and when playing against Watson, they quickly adopt overt DD-seeking behavior.

Multigame wagering model

In most Jeopardy! contests, the winner is determined by performance in a single game. However, the Watson-Jennings-Rutter match utilized point totals over two games to determine the first, second, and third places. This clearly implies that wagering strategies must differ in Games 1 and 2 of the match, and both need to be different from single-game wagering.

Since there is very limited multigame match data available from J! Archive, it would be difficult to model the expected wagering of Jennings and Rutter in the exhibition purely from historical data. Fortunately, we were able to make some educated guesses that considerably simplified the task. First, we predicted that they would wager DDs very aggressively in both games, unless they had an overwhelming lead. This implied that we could continue to use the aggressive heuristic DD model for single games, with a revised definition of what constitutes an "overwhelming" match lead. Second, we also expected them to bet very aggressively in FJ of the first game. This meant that we could treat Game 1 FJ as if it were a DD situation and again use the aggressive heuristic model.

The only situation requiring significant modeling effort was Game 2 FJ. Given limited available match data, only crude estimates could be assigned of the probabilities of various betting strategies. However, it is clear from the data that the wagering of human champions is much more coherent and logical than the observed wagering in regular episodes, and champion wagers frequently satisfy multiple betting constraints. These observations guided our development of revised betting models for Game 2 FJ. As an example, in the case where B has a legal two-thirds bet (suitably defined for matches) and where B can also keep out C, our model for B bets as follows: "bankroll" bet with 35% probability, bet a small random amount that satisfies both the two-thirds and the keepout-C limits with 43% probability, or bet to satisfy the larger of these two limits with 22% probability.

Optimizing Watson strategies using the simulation model

The simulator described previously enables us to estimate Watson's performance for a given set of candidate strategy modules by running extensive contests between a simulation model of Watson and two simulated human opponents. The Watson stochastic-process models use the same performance metrics (i.e., average attempt rate, precision, DD, and FJ accuracy) as in the human models. The parameter values were estimated from J! Archive test sets and were updated numerous times as Watson improved over the course of the project. The Watson model also estimates buzzability, i.e., its likelihood to win the buzz against humans of various ability levels. These estimates were initially based on informal live demo games against IBM Researchers and were subsequently refined based on Watson's performance in the sparring games. We estimated Watson's buzzability at ~80% against average contestants, 73% against Champions, and 70% against Grand Champions.

Computation speed is an important factor in designing strategy modules since wagering, square selection, and buzz-in decisions need to be made in just a few seconds. In addition, the strategy was run on Watson's "front-end," a single server with just a few cores, since its 3,000-core "back-end" was dedicated to QA computations. As a result,

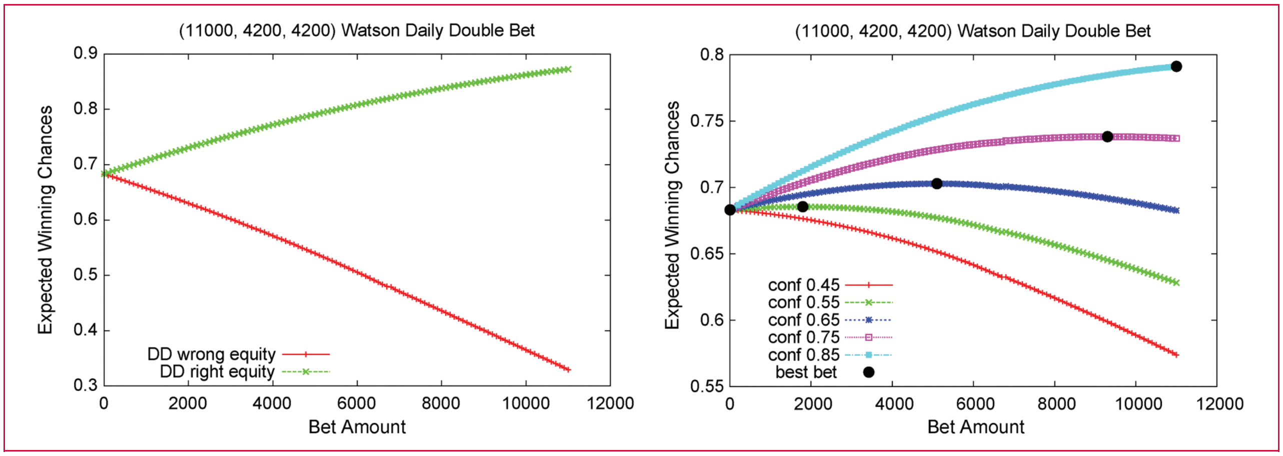

Figure 4

Left: equity estimates getting the DD right (green curve) and wrong (red curve). Right: bet equity curves at five differences in-category confidence levels, from 45% to 85%. Black dots show how the optimal risk-neutral bet increases with confidence.

most of Watson's strategy modules run fast enough so that hundreds of thousands of simulated games can be performed in just a few CPU hours. This provides a solid foundation for evaluating and optimizing the individual strategy components, which are described in the following. Some strategy components (endgame buzz threshold and Game 2 FJ betting) are based on heavyweight Monte Carlo trials; these are too slow to perform extensive offline evaluation. Instead, these strategies perform live online optimization of a single strategy decision in a specific game state.

Daily Double wagering strategy

We implemented a principled approach to DD betting, based on estimating Watson's likelihood of answering the DD question correctly, and estimating how a given bet will impact Watson's overall winning chances if Watson gets the DD right or wrong. The former estimate is provided by an "in-category DD confidence" model. Based on thousands of tests on historical categories containing DDs, the model estimates Watson's DD accuracy given the number of previously seen questions in the category that Watson got right and wrong.

To estimate impact of a bet on winning chances, we follow [5] in using reinforcement learning [6] to train a game-state evaluator (GSE) over the course of millions of simulated Watson-versus-humans games. The GSE consists of a nonlinear function approximator (specifically, a type of artificial neural net called a *Multilayer Perceptron* [7]) that receives a feature-vector description of the current game state and outputs an estimate of the probability that each player will ultimately win the game.

The combination of GSE with in-category confidence enables us to estimate $E(\text{bet})$, i.e., the "equity" (expected winning chances) of a bet, according to

$$E(\text{bet}) = p_{\text{DD}} \times V(S_W + \text{bet}, \ldots) + (1 - p_{\text{DD}}) \times V(S_W - \text{bet}, \ldots), \quad (1)$$

where p_{DD} is the in-category confidence, S_W is Watson's current score, and $V(\,)$ is the game-state evaluation after Watson's score either increases or decreases by the bet and the DD has been removed from the board. We can then obtain an optimal risk-neutral bet by evaluating $E(\text{bet})$ for every legal bet, and selecting the bet with the highest equity. We additionally modified (1) to incorporate known techniques in risk analytics; this achieved a significant reduction in downside risk when getting the DD wrong, at only a slight cost (less than 0.2% per DD bet) in the expected win rate.

Illustrative example

Figure 4 illustrates how the DD bet analysis operates, and how the resulting bet depends on in-category confidence. The example is taken from one of the sparring games, where Watson got four consecutive questions right in the first category at the start of Double Jeopardy! and then found the first DD in attempting to finish the category. At this point, Watson's score was $11,000, and the humans each had $4,200. Watson's in-category confidence took its maximum value of 75%, based on having gotten four out of four correct answers previously in the category. Watson chose to wager $6,700, which is a highly aggressive bet by human standards. (Fortunately, Watson got the DD question right.)

The left figure shows neural net equity estimates for getting the DD right (green curve) and wrong (red curve) at various bet amounts. These curves are extremely smooth with gently decreasing slopes. The right plot shows the resulting equity-versus-bet curve at Watson's actual 75% confidence level (magenta curve), along with four other curves at different confidence values ranging from 45% to 85%. Black dots on each curve indicate the best risk-neutral bet, and we can see how the bet steadily increases with confidence, from $5 at 45%, to approximately $9,300 at the actual 75%, and finally to the entire $11,000 at a (hypothetical) confidence of 85%.

We also note the effect of risk mitigation, which reduced Watson's actual bet from $9,300 to $6,700. According to extensive Monte Carlo analysis of this bet, risk mitigation lowered Watson's equity by only 0.2% (from 76.6% to 76.4%) but reduced downside risk by more than 10% in the event that Watson got the DD wrong.

Multigame DD wagering

As mentioned previously, Games 1 and 2 of the exhibition match required distinct wagering strategies, with both differing from single-game wagering. We trained separate neural networks for Games 1 and 2. The Game 2 net was trained first, using a plausible artificial distribution of Game 1 final scores.

Having trained the Game 2 neural net, we could then estimate the expected probabilities of Watson ending the match in the first, second, or third place, starting from any combination of Game 1 final scores, by extensive offline Monte Carlo simulations. We used this to create three lookup tables, for the cases where Watson ends Game 1 in first, second, or third place, of Watson match equities at various Game 1 final score combinations, ranging from (0, 0, 0) to (72000, 72000, 0) in increments of 6000. (Since adding or subtracting a constant from all Game 1 scores has no effect on match equities, we can without loss of generality subtract a constant so that the lowest Game 1 score is zero.) Since match equities are extremely smooth over these grid points, bilinear interpolation provides a fast and highly accurate evaluation of Game 1 end states. Such lookup tables then enabled fast training of a Game 1 neural net, using simulated matches that only played to the end of Game 1, and then assigned expected match-equity rewards using the table.

An important new issue that we faced in the exhibition match was how to assign relative utilities to finishing in the first, second, and third places. Unlike the sparring games where our sole objective was to finish first, we extensively debated the amount of partial credit that IBM would garner by defeating one of the two greatest Jeopardy! contestants of all time. Ultimately, we decided to base the match DD wagering on full credit for first, half credit for second, and zero credit for a third place finish—such an objective places equal emphasis on finishing first and avoiding finishing third.

Final Jeopardy! wagering

Our approach to FJ wagering involves computation of a Best Response strategy [8] (a standard game-theoretic concept) to the human FJ model presented earlier. We considered attempting to compute a Nash equilibrium [8] strategy but decided against it for two reasons. First, because of the imperfect information in FJ (contestants know their own confidence given the category title but do not know the opponents' confidence), we would in principle need to compute a Bayes-Nash equilibrium, which entails considerably more modeling and computational challenges. Second, it seems far-fetched to assume that Watson's opponents would play their part in a Nash equilibrium since the average contestant has not studied game theory.

Computation of the Best Response proceeds as follows. First, we consult a "Final Jeopardy! prior accuracy" regression model to estimate Watson's confidence given the category title. This model was trained on samples of Watson's performance in thousands of historical FJ categories, using NLP-based feature vector representations of the titles. Second, given Watson's confidence and the human accuracy and correlation parameters, we derive analytic probabilities of the eight possible right or wrong outcomes. Third, for a given FJ score combination, we draw on the order of 10,000 Monte Carlo samples of bets from the human models. Finally, we evaluate the equity of every legal bet, given the human bets and the right/wrong outcome probabilities, and select the bet with highest equity.

After extensive offline analysis, we discovered that the Best Response output could be expressed in terms of a fairly simple set of logical betting rules. As this is much faster than the full Best Response calculation, we deployed a rule-based encapsulation of the Best Response strategy for Watson's sparring game matches. An example betting rule for B stipulates that

> If B has at least two-thirds of A and if B has less than 2C, check whether 2C-B (the amount to cover C's doubled score) is less than or equal to 3B-2A (the maximum two-thirds bet). If so, then bet 2C-B; otherwise, bet everything.

For the exhibition match, we devised live Best Response algorithms for Games 1 and 2 based on Monte Carlo samples of the match human betting models, and probabilities of the eight right/wrong outcomes given Watson's FJ category confidence. For the first-game FJ, we cannot evaluate directly from the FJ outcomes since there is still a second game to play. The evaluation is instead based on interpolation over the lookup tables discussed earlier, denoting Watson's match equities from various first-game score combinations.

For Game 2 FJ, we again devoted substantial effort to interpreting the Best Response output in terms of logical betting rules. The derived rules mostly bet to guarantee a win as A if Watson answers correctly.

For wagering as B, the betting rules would only attempt to finish ahead of A if it did not diminish Watson's chances of finishing ahead of C. This naturally emerged from assigning half-credit for a second place finish in the match utility function. Finally, for wagering as C, the Best Response output was too complex to derive human-interpretable rules; therefore, Watson was prepared to run the live calculation in this case. As it turned out, all of the above work was superfluous since Watson had a lockout in Game 2 of the exhibition match.

Square selection

Selecting a DD can provide an excellent opportunity for a player to significantly boost his game standing while also denying that opportunity to the other players. On the other hand, by gathering information about a category from its low-value questions, contestants can increase their expected accuracy for higher value questions in the same category.

We used the simulator to systematically investigate the relative importance for Watson of the following factors in square selection: finding the DDs, retaining control of the board if a DD is not found, and learning the essence of a category. These studies were performed using Champion and Grand Champion human models, which featured overt DD seeking, aggressive DD wagering, and high DD accuracy. Our results showed that prior to all DDs being revealed, finding DDs is overwhelmingly the top factor in Watson's win rate, and retaining control is second in importance. Learning the essence of a category appears to provide an effective strategy only after all DDs have been found.

These findings led us to deploy an algorithm that selects squares as follows. First, if there are any unrevealed DDs, a square i^* is selected that maximizes $p_{\mathrm{DD}}(i) + \alpha * p_{\mathrm{RC}}(i)$, where $p_{\mathrm{DD}}(i)$ is the probability that square i contains a DD, $p_{\mathrm{RC}}(i)$ is an estimated probability that Watson will retain control of the board if i does not contain a DD, and $\alpha = 0.1$ yielded the best win rate. The first term is calculated using Bayesian inference: we use historic DD frequencies as a prior, combined with evidence from revealed squares according to Bayes' rule, to compute posterior probabilities for the DD locations. The second probability is estimated by combining the simulation model of human performance on regular questions with a model of Watson that adjusts its attempt rate, precision, and buzzability as a function of the number of right/wrong answers previously given in the category. Second, after all DDs in the round have been found, the algorithm switches to selecting the lowest dollar value in the category with the greatest potential for learning about the category: This is based on the number of unrevealed questions in the category and their total dollar value.

Table 2 Simulation win rates versus Grand Champions using various square selection strategies.

Square selection strategy	*Win probability* (200,000 trials)
Top-to-bottom in best column	0.611
Bayesian DD	0.677
Bayesian DD, post-DD learning	0.680
Bayesian DD + 0.1 p_{RC}, post-DD learning	0.681

Relative contributions of the three factors can be seen in **Table 2**. We first measured Watson's win rate against simulated Grand Champions using a baseline strategy of always selecting squares in the column with the highest estimated accuracy in a top-to-bottom order. We then find an improvement of 6.6% by switching to Bayesian DD seeking if DDs are available. We then obtain a further 0.3% improvement by using the previously described learning strategy with no remaining DDs. Finally, we obtain 0.1% gain by including $\alpha * p_{\mathrm{RC}}(i)$ per above with $\alpha = 0.1$.

Confidence threshold for attempting to buzz

Watson attempts to buzz in if the confidence in its top-rated answer exceeds an adjustable threshold value. In most game states, the threshold is set to a default value that is tuned to maximize expected earnings. Near the end of the game, the threshold may vary significantly from this default value. One special-case policy that we devised for endgames uses a "lockout-preserving" calculation. For Round 2 states with no remaining DDs, if Watson has a big lead, we calculate whether Watson has a guaranteed lockout by not buzzing on the current square. If so, and if the lockout is no longer guaranteed if Watson buzzes and is wrong, we prohibit Watson from buzzing, regardless of confidence.

In principle, there is an exact optimal buzz-in policy $[B_0{}^*(c,D), B_1{}^*(c,D), B_2{}^*(c,D), B_3{}^*(c,D)]$ for any game state with a question in play, given Watson's confidence c and the dollar value D of the current question. The policy is a four-component vector as there are four possible states in which Watson may buzz: the initial state, the first rebound where human #1 answered incorrectly, the first rebound where human #2 answered incorrectly, and the second rebound where both humans answered incorrectly. The optimal policy can be calculated using Dynamic Programming (DP) techniques [9]. This involves writing a recursion relation between the value of a current game state with K questions remaining before FJ and values of the possible successor states with $K-1$ questions remaining as follows:

$$V_k(s) = \int \rho(c) \sum_{j=1}^{5} p(D_j) \max_{\vec{B}(c,D_j)} \times \sum_{\delta} p(\delta|\vec{B},c) V_{k-1}\left(s'\left(\delta, D_j\right)\right) dc, \qquad (2)$$

where $\rho(c)$ is the probability density of Watson's confidence, $p(D_j)$ denotes the probability that the next square selected will be in row j with dollar value $D_j = 400 * j$, the max operates over Watson's possible buzz/no-buzz decisions, $p(\delta|B,c)$ denotes the probability of various unit score-change combinations δ, and s' denotes various possible successor states after the D_j square has been played, and a score change combination δ occurred.

Exact DP computation of the optimal buzz-in policy entails expanding the root state to all possible successor states, going all the way to FJ states, evaluating the FJ states by Monte Carlo trials, and then working backwards using (2) to ultimately compute the optimal buzz policy in the root state. However, this is generally too slow to use in live play, where the buzz-in decision must take at most ~1 to 2 seconds. We therefore implemented an Approximate DP calculation in which (2) is only used in the first step to evaluate V_K in terms of V_{K-1}, and the V_{K-1} values are then based on plain Monte Carlo trials [10–12]. Due to slowness of the exact calculation, we were unable to estimate accuracy of the approximate method for $K > 5$. However, we did verify that Approximate DP usually gave quite good threshold estimates (within ~5% of the exact value) for $K \leq 5$ remaining squares; therefore this was our switchover point to invoke Approximate DP as deployed in the live sparring games.

Our buzz-in algorithm easily handles, for example, a so-called "desperation buzz" on the last question, where Watson must buzz and answer correctly to avoid being locked out. From time to time, it also generates spectacular movements of the buzz threshold that are hard to believe on first glance but that can be appreciated after detailed analysis. An example taken from the sparring games is a last-question situation where Watson has $28,000, the humans have $13,500 and $12,800, and the question value is $800. The (initially) surprising result is that the optimal buzz threshold drops all the way to zero. This is because after buzzing and answering incorrectly, Watson is no worse off than after not buzzing. In either case, the human B player must buzz and answer correctly in order to avoid the lockout. On the other hand, buzzing and answering correctly secures the win for Watson; therefore, this is a risk-free chance to try to buzz and win the game.

Summary of performance evaluation

We have performed extensive analysis to assess the faithfulness of the simulator in predicting live game results and to measure the performance benefits of our advanced strategies, over both simpler baseline strategies as well as strategies used by humans. Detailed documentation of these analyses will be presented in a more extensive future publication. In brief, the simulator matched live results in the sparring games within sampling error according to every metric that we examined. These include quantities such as the following: Watson's win rate, Watson and human's lockout rates, Watson's rate of leading going into FJ, Watson and human's average scores, and Watson's average board control. The simulator also accurately predicts the percentage of DDs found by Watson and the average time it takes to find the first, second, and third DD in each game.

In comparing each strategy module with a corresponding baseline heuristic, we obtained the following results: The improvement due to advanced DD betting was ~6.5% more wins, whereas the square selection improvement ranged up to 7.6%, as we discussed previously. For FJ betting, we estimate a ~3% gain over a heuristic algorithm that always bets just enough to close out when leading and that bets everything when trailing. We have no data on heuristic endgame buzzing, but a conservative guess is that our Approximate DP buzzing would achieve ~0.5% to 1% greater win rate. The cumulative benefit of these gains appears to be additive, as the simulator estimates Watson's win rate at 50% using all baseline strategies, versus 70% using all advanced strategies.

Watson's ability to find DDs more effectively than humans can be seen from the fact that its rate of finding DDs exceeds its average board control. This was true not only in simulation but also in live sparring games, where Watson found 53.3% of the DDs but only had 50.0% average board control. We can also show superiority of Watson's wagering strategies by analysis of human bets in the J! Archive data. For example, the historic win rates in nonlocked FJ are 65.3% as A, 28.2% as B, and 7.5% as C. Using historic replacement of the recorded human bets with Watson's Best Response bets, these win rates increase to 67.0%, 34.4%, and 10.5%, respectively. We have also performed extensive Monte Carlo analysis of human DD bets on the last DD of the game. The results show that the average equity loss of human DD bets is between 3.2% and 4.2% per bet, which is an order of magnitude worse than Watson's loss rate of 0.25% per bet.

Conclusion

We have presented an original quantitative approach to strategy for playing the television game show *Jeopardy!*. Our approach is comprehensive, covering all aspects of game strategy: wagering, square selection, attempting to buzz-in, and modifications for multigame tournament matches. A key ingredient in this paper was the development of an original simulation model of the game and its human contestants.

Apart from numeric metrics showing outperformance of human strategies, it is plainly evident that our strategy algorithms achieve a level of quantitative precision and real-time performance that exceeds human capabilities. This is particularly true in the cases of DD wagering and endgame buzzing, where humans simply cannot come close to matching the precise equity and confidence estimates and complex decision calculations performed by Watson in real time. Nor are humans capable of performing live Bayesian inference calculations of DD location likelihood. As such, the software that we have developed could prove to be a valuable teaching tool for prospective contestants. If it were made widely accessible and reoriented to optimizing human strategies, this could result in broad improvements in strategic decision-making as seen on the show. Such improvements have already occurred in numerous games such as Chess, Checkers, Othello, and Backgammon, after the best programs surpassed the top humans, and we expect Jeopardy! to be no different in this regard. We are investigating deployment of calculators of DD wagers and DD location probabilities on J! Archive. This would nicely complement the existing FJ wager calculator and make our methods widely accessible for study by prospective contestants.

Looking beyond the immediate Jeopardy! domain, we also foresee more general applicability of our high-level approach to coupling decision analytics to QA analytics, which consists of building a simulation model of a domain (including other agents in the domain), simulating short-term and long-term risks and rewards of QA-based decisions, and then applying learning, optimization, and risk analytics techniques to develop effective decision policies. We are currently investigating applications of this high-level approach in health care, dynamic pricing, and security (i.e., counter-terrorism) domains.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

**Trademark, service mark, or registered trademark of Jeopardy Productions, Inc., in the United States, other countries, or both.

References

1. Jeopardy! Gameplay. [Online]. Available: http://en.wikipedia.org/wiki/Jeopardy!#Gameplay
2. D. Billings, A. Davidson, J. Schaeffer, and D. Szafron, “The challenge of poker,” *Artif. Intell.*, vol. 134, no. 1/2, pp. 201–240, 2002.
3. J! Archive. [Online]. Available: http://www.j-archive.com
4. F. Leisch, A. Weingessel, and K. Hornik, “On the generation of correlated artificial binary data,” Vienna Univ. Econ. Bus. Admin., Vienna, Austria, Working Paper 13, 1998.
5. G. Tesauro, “Temporal difference learning and TD-Gammon,” *Commun. ACM*, vol. 38, no. 3, pp. 58–68, Mar. 1995.
6. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
7. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1987, pp. 318–362, PDP Research Group.
8. D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.
9. D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 1995.
10. G. Tesauro and G. R. Galperin, “On-line policy improvement using Monte-Carlo search,” in *Advances in Neural Information Processing Systems*, M. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 1068–1074.
11. B. Sheppard, “World-championship-caliber scrabble,” *Artif. Intell.*, vol. 134, no. 1/2, pp. 241–275, Jan. 2002.
12. M. Ginsberg, “GIB: Steps toward an expert-level bridge playing program,” in *Proc. IJCAI*, 1999, vol. 16, pp. 584–593.

Received July 18, 2011; accepted for publication January 18, 2012

Gerald Tesauro *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (gtesauro@us.ibm.com).* Dr. Tesauro is a Research Staff Member at the T. J. Watson Research Center. He is best known for developing TD-Gammon, a self-teaching neural network that learned to play backgammon at the human world championship level. He has also worked on theoretical and applied machine learning in a wide variety of other settings, including multi-agent learning, dimensionality reduction, computer virus recognition, computer chess (DEEP BLUE*), intelligent e-commerce agents, and autonomic computing. Dr. Tesauro received B.S. and Ph.D. degrees in physics from University of Maryland and Princeton University, respectively.

David C. Gondek *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (dgondek@us.ibm.com).* Dr. Gondek is a Research Staff Member and Manager at the T. J. Watson Research Center. He received a B.A. degree in mathematics and computer science from Dartmouth College in 1998 and a Ph.D. degree in computer science from Brown University in 2005. He subsequently joined IBM, where he worked on the IBM Watson Jeopardy! challenge and now leads the Knowledge Capture and Learning Group in the Semantic Analysis and Integration Department.

Jonathan Lenchner *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (lenchner@us.ibm.com).* Dr. Lenchner is a Senior Technical Staff Member at the T. J. Watson Research Center. He created the framework into which the various Jeopardy! strategy components were integrated and tested using tens of millions of simulated Jeopardy! games. Always a student of games of strategy, he is a former chess master and wrote a strong chess-playing program while a sophomore at Dartmouth College. His main work at IBM is to invent ways to make data centers and other facilities more energy efficient. Together with his colleagues, he has recently built a robot for mapping and thermally monitoring previously unseen data centers.

James Fan *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (fanj@us.ibm.com).* Dr. Fan is a Research Staff Member in the Semantic Analysis and Integration Department at the T. J. Watson Research Center, Yorktown Heights, NY. He joined IBM after receiving his Ph.D. degree at the University of Texas at Austin in 2006. He is a member of the DeepQA Team that developed the Watson question-answering system, which defeated the two best human players on the quiz show Jeopardy!.

Dr. Fan is author or coauthor of dozens of technical papers on subjects of knowledge representation, reasoning, natural-language processing, and machine learning. He is a member of Association for Computational Linguistics.

John M. Prager *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (jprager@us.ibm.com).* Dr. Prager has been working in technical fields related directly or indirectly to question answering for most of his professional career. Most recently, while at the T. J. Watson Research Center, he has been part of the Watson project, building a system that plays the *Jeopardy!* quiz-show game. He has been involved in both the algorithms area, in particular working on puns and other wordplay, and the strategy area. Previously, he led IBM's successful entries in Text REtrieval Conference Question-Answering (TREC-QA) tasks, an annual evaluation at the National Institute of Standards and Technology (NIST). Prior to that, he worked in various areas of search, including language identification, web search, and categorization. He has contributed components to the IBM Intelligent Miner for Text product. For a while in the early 1990s, he ran the search service on www.ibm.com. While at the IBM Cambridge Scientific Center, Cambridge, Massachusetts, he was the project leader of the Real-time Explanation and Suggestion project, which would provide users with help by taking natural-language questions and processing them with an inference engine tied to a large repository of facts and rules about network-wide resources. He has degrees in mathematics and computer science from the University of Cambridge and in artificial intelligence from the University of Massachusetts; his publications include conference and journal papers, nine patents, and a book on Alan Turing.