# Detecting strategic moves in HearthStone matches

Boris Doux[1], Clement Gautrais[2], and Benjamin Negrevergne[1]

[1] Inria Rennes – LACODAM Team,
`firstane.lastname@inria.fr`
[2] IRISA / University of Rennes I
`firstane.lastname@irisa.fr`

**Abstract.** In this paper, we demonstrate how to extract strategic knowledge from gaming data collected among players of the popular video game *HearthStone*. Our methodology is as follows. First we train a series of classifiers to predict the outcome of the game during a match, then we demonstrate how to spot key strategic events by tracking sudden changes in the classifier prediction. This methodology is applied to a large collection of HeathStone matches that we have collected from top ranked European players. Expert analysis shows that the events identified with this approach are both important and easy to interpret with the corresponding data.

**Keywords:** e-sports analytics, descriptive analysis

## 1 Introduction

HearthStone (HS) is a popular online video game in which players fight one another using virtual playing cards. To defeat their opponent, players have access to a large number of cards, that they can use to trigger a variety of offensive or defensive moves. The cards in HS have complex synergies which are exploited by most experienced players to set up powerful strategies. Over the time, experienced players have accumulated a large body of strategic knowledge about the combinations of cards and how to use them. However, this knowledge remains mostly inaccessible to novice players.

In this paper we are interested in identifying key strategic events during matches to help novice players to decode and learn from matches between skillful players. The problem of extracting interesting strategic insights from game data has been addressed in many different ways which typically compromise between quality of the knowledge extracted, and the amount of external expert knowledge that needs to be used to extract them. For example Bosc et al. [1] use pattern mining techniques to identify strategies from StarCraft replay data. However this approach requires expert analysis of a large number of patterns which can be tedious. To this day, what really is a strategic insight remains an open question.

In general, what is truly interesting is a subjective matter. However, many games, including HearthStone, have been designed to promote small events with

big impact on the outcome of the game. For example in HS, playing a particular card at a particular time can have a dramatic impact on the rest of the match, and thus it should be considered as a strategic event of the ongoing match.

Building on this observation, we explore the hypothesis that a strategic event in a game is one that impacts the outcome of the game. To validate this hypothesis, we train a series of classifiers at different stages of a match and we analyze the events which are associated with a change in the classifier prediction over time.

To identify truly interesting strategic events, it is crucial to obtain reasonable classifier confidence. But predicting the outcome of a match with confidence using only features observed during a match (such as players health points or number of creature) is a difficult task even for experienced human experts, and *a fortiori* for simple learning algorithms. To overcome this difficulty, we focus on the end of the match, where the winner is easy to detect from simple game features such as players healtpoints, and step backward in the match turns, until the classifier confidence is too low. As we will show in our experiment, this technique enables us to spot a number of interesting strategic events up to 8 turns before the final turn.

We apply this approach on a large collection of match data which we have collected from real players. Our experiments show that this approach can be used to identify events which are particularly relevant and further analysis of the classifiers parameters, such as feature weights, can be used to better understand these events.

In summary, this paper contributes in three different ways. First we introduce a new dataset[1] which includes over two thousands HearthStone matches played by skillful players (among the top 1% European players). Second we demonstrate how to train classifiers which are sufficiently accurate to detect events impacting the outcome of the game. Finally, we demonstrate how to identify strategic events using changes in classifiers prediction and we derive number of strategic insights from the analysis of the classifiers. These results validate our initial hypothesis that events impacting the outcome of the match can provide strategic insights. As we will show, these insights are not specific to one particular match and provide useful knowledge for novice players.

## 2 Related works

A large body of work have already been dedicated to the problem of sport analytics for various purposes. For example, predicting the outcome of the match is a popular problem due to its connections with betting. Match outcome prediction can be done either *offline*, (for example see the work by Goddard et al. [3] applied to football), or *online* (for example, the work by Klaassen et al. [5] applied to Tennis).

Compared to traditional sports, e-sports offer an interesting test-bed for novel match data analysis techniques because the data is collected in a controled envi-

---

[1] The dataset is available at https://bitbucket.org/Valnora/hsdataset

ronment (the game engine) but still reflects actions performed by highly competitive human players (see [10] and [4] for discussion on e-sports). Thanks to this favorable setting many researcher such as Rioult [8], Bosc [1] and Lewis [6] and their colleagues, have used techniques, such as pattern mining and topological analysis, to identify winning game strategies.

Finding highlights in match data is another interesting problem which can be used to build game summaries and to help the analysis of raw match data. Most work in the literature involve the prior definition of game events and highlights by an expert [2]. These game events can be detected through video analysis, or from spectators reaction [11][7].

Our work differs from these, as we want to extract key events, solely based on game features, rather than external information, such as expert knowledge or crowd cheering.

E-sports is being more and more studied in the research community, as match data are usually easily available. Indeed, many popular electronic games have a replay feature that record game actions. These replays can then be analyzed, without having to use external websites game summaries, as it is usually the case for traditional sports. Nevertheless, HearthStone does not have this replay feature and game data have to be extracted using custom tools, leading to the fact that there is currently no publicly available detailed data of HS matches. Our work is interesting as only few work have been conducted on the game HearthStone, and their main concern is to build an Artificial Intelligence that is able to compete with other, scripted or random, artificial players [9]. Our approach is, to the best of our knowledge, the first one that aims at providing strategic insights about the game, from real players data.

## 3   HearthStone basics

HearthStone, is an online card game with a medieval fantastic flavor. Matches in HS involve two players, playing in turns, one at a time. Each turn starts with one player drawing a card from her deck, and then playing a small sequence of actions using the cards in her hand. When the player has performed all the actions that she wants to play, the turn ends, and the other player can start playing. Players in HS are incarnated by a character who has a fixed number of *health points* and a unique ability. In order to win a match, a player has to bring the health points of the opponent's character down to zero.

Cards can be used to perform a variety of actions but the majority of them invoke magic creatures which have attack points and health points (the health points of the creature are independent from the player's health points). The creatures that have been invoked can then be used to attack the opponent creatures, or the opponent character directly.

Before they can play any match, players are required to build a *deck*, which is a small set of cards they want to play with. Because cards can have important synergies between them, all decks are not equally powerful, and the best decks enable the player to draw powerful combinations of cards during a match with

high probability. Furthermore, since each character class has access to an extra set of cards, players have to build decks which can deal with a large spectrum of opponents. As a consequence, creating novel powerful decks requires lots of expertise and has become an important part of the game.

## 4 The HearthStone dataset

To build strategic knowledge about HearthStone and help novice players, we have collected match data from several skilled HearthStone players using a *game tracker*[2]. The tracker runs in the game client of one player and records all the data that this player can see. We track three players who have ranks ranging from rank 10 to legend (good players to elite players). The opponents are automatically selected by the game engine at the beginning of each match and since the game engine uses players ranking to ensure balanced matches, the opponent players in our dataset have similar levels as the tracked players.

Each record in the dataset $D$ is a pair $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i$ is a list of feature vectors $\mathbf{x}_i^1, \ldots, \mathbf{x}_i^n$ describing each one of the $n$ turns of the corresponding match $i$, and where $y_i$ is a boolean label indicating whether the tracked player won the match or not. A feature vector $\mathbf{x}_i^k$ describes a turn $k$ using the following features: *turn id* (integer), *cards played during the turn* (list of cards), *number of cards in player hand* (integer), *number of cards in the opponent hand* (integer), *player's creatures on the board* (list of cards), *opponent's creature on the board* (list of cards), *player's turn or opponent turn* (boolean), *player's health* (integer) *opponent's health* (integer) , *player's armor* (integer), *opponent's armor* (integer). In the following, we denote $D^k$ the dataset restricted to the feature vector for the turns $k$ only. (i.e. $D^k = \{(\mathbf{x}_i^k, y_i) : \forall (\mathbf{x}_i, y_i) \in D\}$). In our dataset the turns are numbered backward starting from the final one (i.e. turn 0 is the final turn). This choice is discussed in the following section.

## 5 Extracting strategic knowledge from the HearthStone dataset using classifiers

Our hypothesis is that a strategic event is one that has an important impact on the outcome a match. In other words, if a particular event (such as a playing a card) drastically changes the estimated winner prior to this event, we consider it to be a strategic event.

To validate this hypothesis using the data collected as described in Section 4, we apply the following methodology.

(i) We train a sequence of classifiers $\langle f_1, \ldots, f_n \rangle$ to estimate the winning player (i.e. the class label $y$) after each turn of a match. To do so, we train each classifier $f_k$ using the match data collected at turn $k$ (i.e. using the restricted dataset $D^k$ as introduced in Section 4). Remark that training a sequence of classifiers is preferred over training a single classifier because features do not

---

[2] Tracker available at https://github.com/HearthSim/Hearthstone-Deck-Tracker

have the same importance at different turns. (This effect is demonstrated later in Figure 3.)

(ii) We use the sequence of classifiers to make a sequence of predictions about the winner after each turn. Given a match description $\mathbf{x}$, the sequence of classifiers will produce a sequence of predictions $\langle f_1(\mathbf{x}^1), \ldots, f_n(\mathbf{x}^n) \rangle$. In our experiments, we use probabilistic classifiers so that $f_k(\mathbf{x}^k)$ is the estimated probability of a win given the game state at turn $k$ (i.e. $f_k(\mathbf{x}) = \Pr(y = 1|\mathbf{x}^k)$).

(iii) We select few match descriptions $\mathbf{x}$ having the largest difference between two consecutive predictions $|f_k(\mathbf{x}^k) - f_{k+1}(\mathbf{x}^{k+1})|$.

(iv) We analyze each corresponding match in the light of the feature coefficients, and show that changes in the classifiers predictions are indeed correlated with highly strategic events in the match.

The main difficulty is to achieve sufficient classifier accuracy to obtain significant results. Several papers have demonstrated how to achieve good accuracy in the context of game outcome prediction using standard machine learning algorithms (for example Goddard et al. used regression techniques to forecast football match result [3]), but the predictions rely on player statistics aggregated over historical (e.g. win/loose ratio). These statistics do not change over the course of a match and cannot be used to make varying predictions as required by our approach. In contrast, classifiers based on game features only, often make less reliable predictions which cannot be used to produce significant results.

To overcome this problem the key observation is that predicting the outcome of the game becomes easier as we approach the end of the game: predicting the winner after the final round is (almost) trivial and becomes increasingly difficult as we approach the beginning of the game. Building on this observation, we have numbered the turn starting from the final one (as described in the previous section) and we focus on the final turns, where the classifier confidence is high enough. As we will show in our experiments we are able to achieve reasonable classifier accuracy up to 8 turns before the final turn.

## 6 Experiments

In this section we demonstrate the applicability of our approach on the HearthStone dataset. In particular, we first look at the classifier accuracy at different stages of the game **(Q1)**. Then, we validate experimentally that changes in the classifier prediction are correlated with key strategic events **(Q2)**. Finally, we demonstrate important strategical insights about the key events that can be derived from the analysis of the corresponding match data **(Q3)**.

*How reliable are classifiers at predicting the winner n turns before the final turn (Q1)* In our approach, the classifier accuracy is critical to derive useful knowledge, thus we start this experimental section by evaluating the accuracy of the classifiers at predicting the outcome of the game. Only the most accurate classifiers will be used to validate the subsequent experiments (**Q2**) and (**Q3**).

We train a series of classifiers (one for each turn) as described in Section 5. The classifiers are trained using three different learning algorithms: Naive Bayes (NB) and Logistic Regression Classification (LR), which produce easy to interpret models based on linear predictors (required for **Q3**), and Random Forest Classification (RF) which will serve as a comparison for non-linear models.

The classifiers accuracy and f-score for the 10 final turns are presented in Figure 1 (left). On these plots, the final turn is labeled 0 on the x-axis. Unsurprisingly, the accuracy tops at the final turn (where the winner is already set) and decreases as we get closer to the beginning of the match. The predictions remain significantly better than the majority class prediction (represented by the horizontal line in Figure 1) up to 8 turns before the end of the game, which is a good result. Remark that the accuracy in the last turn is not 100% because many players choose to concede when they are in a bad shape. In this case, predicting the winner at the last turn remains non-trivial and the classifiers can still make errors.

If we compare the performances of the different training algorithms, we can see that classifiers trained with LR achieve the best results overall. RF occasionally does better, but mostly on the last turn which is not important for the reason mentioned in the previous paragraph.

The instability of the classifiers accuracy on the left hand side is an artifact of our decision to number the turns starting from the final one. Even numbered turns are played by the player that will eventually win the match whereas odd numbered turns are played by the player that will be defeated. At the end of the game, turns played by the winner will often increase the gap between players' health points, thus making the classifier more confident in its prediction. On the other hand, turns played by the loosing opponent will often reduce the gap (for example, the loosing player can choose to heal himself) making the classifier less confident in its prediction.

In order to balance this undesirable effect, we train classifiers on *double turns*. A double turn include all the actions played by the player at the turn $n$, and all the actions played by the opponent at turn $n + 1$. As one can see, on the right hand side in Figure 1, predictions made using this method are both more accurate and more stable over time.

In the light of these results, we based subsequent experiments on classifiers trained with LR on the last 8 turns, using double turns.

*Are changes in classifier confidence correlated with important events (Q2) and can we derive strategical moves from classifiers analysis (Q3)* Our goal is to validate our main hypothesis, that a strategical event in a match is associated with a change in the classifier confidence. To do this, we first plot the classifier confidence in predicting the winner at different stages of the game for a variety of manually selected matches which exhibit a sudden change in the final turns. We discuss three of these games and demonstrate how to derive strategical insights from the analysis of the corresponding data (Q3).
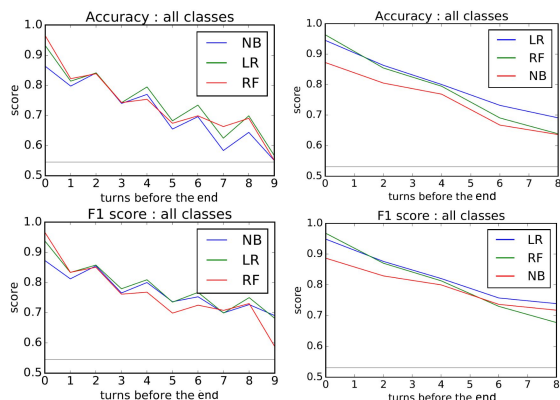
**Fig. 1.** Variation of the classifiers performance as we approach the final turn (turn 0). Measured by accuracy score (top) and by f1 score (bottom).
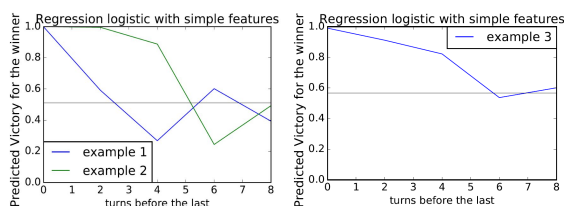


**Fig. 2.** Variation of the classifiers confidence over time for 3 selected matches.

The first game, (Example 1 in Figure 2), is significant because the classifier predicts the incorrect winner with relatively high confidence only 4 turns before the final turn. At this turn, the final winner does not have any creature on the board. Without creatures players can deal a very limited amount damage to their opponent, and thus the situation looks more favorable to the other player. In this situation, it is more reasonable for the classifiers and for novice players to predict that the player with the most creatures will win. However, this experienced player chose to play a rare combination of three cards which allowed him to put many weak creatures that deal damage to the opponent, whereas dealing damage with a creature usually requires an extra turn. This strategy has lead the player to reverse the course of the match and win within the next turns.

In the second game, the classifier also predicts the wrong winner 6 turns before then end of the match. Nevertheless, the prediction greatly changes during the next turns. Between turn 4 and 5, the player choose to play a particular card called *Flamestrike*. The community of experienced HearthStone players has identified this card as one of the most powerful cards to play with a mage character. This card deals important damages to all creatures, and opponents
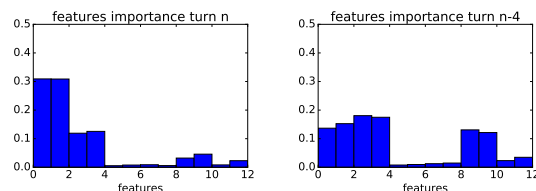
**Fig. 3.** Features weights at different stages of the game.

should not play their important creatures before this card is played. However, inexperienced players are often tricked by it.

In Figure 2 (right) an important increase of the classifier confidence occurs at turn 4. Further analysis of this match shows that, similarly to the previous example, a powerful card was played. This card is called *Savannah Highmane*, and has important synergy with the Hunter class. As we can see here, playing this card gave an important edge to the winning player.

The previous analyses were performed on single games and identify precise strategic moves which lead to an important change in the classifier's confidence. By looking at the corresponding match data, we were able to confirm that such events, are strongly correlated with important strategic moves. In some cases, this analysis corroborates expert knowledge.

Important changes in the classifiers confidence can also be analyzed through the changes in features importance of classifiers at different stages of the game. For example, in the last turns of Example 1, the future loser's character loses many life points. As one can see, the classifiers for the last turns pays more attention to life points ((features 0 and 1 in Figure 3 left) than classifiers at early stages of the game ((features 0 and 1 in Figure 3 right). The opposite effect is observed for the number of cards (features 8 and 9 in Figure 3), whose impact on the predicted outcome decreases as the game progresses. This explains the fact that, even though the future winner has less cards than his opponent in the last turns, his higher health leads to a high probability of victory.

To conclude, we are able to identify strategic moves that can corroborate current community knowledge. Moreover, through the analysis of classifiers features, we are able to find strategic resources (health, cards or creatures) at each stage of the game.

## 7 Conclusion

In this paper, we have demonstrated how to use standard classification techniques to identify key strategical events in HearthStone. Our underlying hypothesis is that strategical events in HearthStone are associated with a change in the classifier prediction. In order to validate this hypothesis, we have collected a large dataset of HearthStone matches from experienced players. Then, we proposed a technique to train accurate classifiers to spot important strategical events in the dataset. We have then shown that the analysis of these events can

provide strategical insights which often corroborate expert knowledge available. Such strategical knowledge would have been tedious to derive from the entire dataset.

Although the applicability of our approach was demonstrated on HS only, we have not made use of any specificity of the game. Therefore this method can also be aplied to other e-sports with a similar game structure (i.e. two players, turn based). Extending this approach to traditional sports is also an interesting challenge. Just like HS, many sports are two players games and are also turn-based to some extend. For example, in Tennis — as well as in many other net games — each player is given a limited amount of time to control the ball and take advantage of the situation. Automatic analysis of Tennis matches is significantly more challenging than HS matches, since both time and moves are continuous (rather than discrete in HS), but results in this domain would be very valuable for both players, and for spectators.

## 8   Acknowledgment

## References

1. Guillaume Bosc, Mehdi Kaytoue, Chedy Raıssi, and Jean-François Boulicaut. Strategic pattern discovery in rts-games for e-sport with sequential pattern mining. Technical report, RR-LIRIS-2013-012, 2013.
2. Peng Chang, Mei Han, and Yihong Gong. Extract highlights from baseball game video with hidden markov models. In *Int. conf. on Image Processing*, 2002.
3. John Goddard and Ioannis Asimakopoulos. Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1):51–66, 2004.
4. Juho Hamari and Max Sjöblom. What is esports and why do people watch it? *Available at SSRN 2686182*, 2015.
5. Franc JGM Klaassen and Jan R Magnus. Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148(2):257–267, 2003.
6. Joshua M Lewis, Patrick Trinh, and David Kirsh. A corpus analysis of strategy video game play in starcraft: Brood war. In *Annual conf. of the cognitive science society*, 2011.
7. Baoxin Li and M Ibrahim Sezan. Event detection and summarization in sports video. In *Content-Based Access of Image and Video Libraries, 2001.(CBAIVL 2001). IEEE Workshop on*, 2001.
8. François Rioult, Jean-Philippe Métivier, Boris Helleu, Nicolas Scelles, and Christophe Durand. Mining tracks of competitive video games. *AASRI Procedia*, 8:82–87, 2014.
9. David Taralla, Zixiao Qiu, Antonio Sutera, Raphaël Fonteneau, and Damien Ernst. Decision making from confidence measurement on the reward growth using supervised learning: A study intended for large-scale video games. In *Int. Conf. on Agents and Artificial Intelligence*, 2016.

X

10. Michael G Wagner. On the scientific relevance of esports. In *International Conference on Internet Computing*, pages 437–442, 2006.
11. Zengkai Wang, Junqing Yu, Yunfeng He, and Tao Guan. Affection arousal based highlight extraction for soccer video. *Multimedia Tools and Applications*, 73(1):519–546, 2014.