Data mining of chess chunks: a novel distance-based structure

A. Linhares
The Getulio Vargas Foundation, Brazil

Abstract

What can the disciplines of artificial intelligence and the cognitive sciences obtain from the application of data mining methodology? There is a great chasm between traditional artificial intelligence methods and real human cognition. The notorious game of chess is a good example, having been traditionally modeled by search engines with advanced pruning techniques. There is, however, considerable psychological evidence that points to the view that human players do not carry out such types of processes. A more suitable alternative view seems to be that chess masters are carrying out a form of mining, carefully categorizing each chunk perceived in a position and gradually building a complex dynamic structure to represent the pressures embedded in each position. This is not only a new application area for data mining, but it is also of utmost importance to psychological theories to model the chess game as a mining process. But what type of data preparation is required? In this paper a chess representation structure, referred to as a distance graph, is presented and discussed. This representation seems to account for numerous characteristics of human player’s psychology, such as: (i) the movement of eye saccades between related pieces, (ii) the carefully counting of movement combinations in selected strategic points in the game, (iii) the presence of empty spaces in chess player’s chunks and (iv) a much more selective search than the classical exhaustive tree model. The psychological plausibility of the computational model opens a new avenue of research for data mining methodology as applied to the cognitive sciences and to testing new theories of artificial intelligence.
1 Introduction

Data mining has often been used in business applications and in engineering problems, but it can also be used to generate basic original results in physics, medicine, and other sciences [1]. In this paper we will be concerned with understanding the following question: what can the disciplines of artificial intelligence and the cognitive sciences obtain from the application of data mining methodology? Towards this, we will consider the notorious problem of chess, one of the most widely studied problems in artificial intelligence and in cognitive psychology. Our intention is twofold: initially we intend to show that data mining methodology may provide the crucial tool for analysis of a hypothesis in this arena. Moreover, we launch a crucial data preparation research agenda for this data-mining based computer chess.

Figure 1: How can the intuitive understanding of a position arise? In this position, chess masters are able to immediately perceive some important features which lead to a winning strategy for white.

Chess has been traditionally viewed in artificial intelligence as a classic symbolic-paradigm search problem: given an initial state, and an overarching goal, the system searches through a myriad of candidate options in search of an optimum. Deepblue, the paramount example, is said to have achieved astonishing speeds up to 330 million positions evaluated per second on the match against Kasparov [2]. However, as psychologists have repeatedly pointed out, humans do not play chess by brute-force search – in fact, it is well established that skilled human players hardly ‘search trough’ that great number of nodes of the game tree, and upper bound estimations of this number reach at most 100 moves for each position [3-12]. It is understood that the key to understanding human chess capability lies not in search (or inference capacity) but really in the underlying subcognitive processes of knowledge acquisition and perception.

World chess champion Cuban José Raul Capablanca once remarked of his personal, subjective, experience: “I know at sight what a position contains. What could happen? What is going to happen? You figure it out, I know it!”. In another occasion, talking about the numerous possibilities that less-skilled players usually considered at each board position, he bluntly remarked: “I see only one move: The best one.” [13]. This process of ‘just knowing’, or of ‘seeing only the best movement’, is the hallmark of human intuition and has been
incredibly elusive to computational models. There is a great chasm between artificial intelligence methods and real human cognition [14,15].

Wilkins [16] pointed out a perfect example of this chasm. Take the board in figure 1, for instance. An experienced player will instantly recognize some of its broad patterns, which naturally lead to the solution. Note the following patterns: (i) the pawn chains block each other, and the white pawn on F6 is the only one capable of moving; (ii) the white pawn on F6 cannot upgrade as long as the king remains in such a short distance; (iii) The black king cannot move away from the pawn on F6 and thus cannot attack the white pawn structure at its unprotected end; (iv) the blocking pawn chains divide the board in two sides, the only passage between them lying on the queen rook file. Having noticed these patterns, the solution becomes clear: white has a winning sequence by moving across the queen rook file and either forcing the black king to respond to a movement from the pawn at F6 or to distance itself from the white king, and either way permitting white to safely upgrade at least one of its pawns.

After these broad patterns are perceived, expectations are formed, and even inexperienced players can intuitively understand such lines of reasoning. Computer chess, in its classical form at least, gets lost in the combinatorial explosion and thus has great difficulty with this position. Wilkins [16] has noted that, “since white’s advantage would take more than 20 ply to show up in most evaluation functions used by computer programs”, these programs should probably decide to move the white king to a more centered position. An experiment with the ‘solve for mate’ function of the Chessmaster 7000 program, for instance, confirms this point. After laboring for days, over three billion movement combinations were considered by the program without finding the right course of action. By now it should be clear that a program capable of using the concepts mentioned above would not even consider moving the white king to the center of the board. Let us discuss how such a problem may be handled by an intuitive chess machine, a problem which perhaps has been best stated by Atkinson [13]:

“The master’s superior play is due to ‘sense of position’, an intuitive form of knowledge gained from experiencing a great variety of chess situations. Intuitive knowledge is perhaps the most important component of expertise in any human discipline, yet intuition has been one of our least understood phenomena. Intuition is an automatic, unconscious part of every act of perception. It is often associated with emotion: an intuitive realization can ‘feel right’ and evoke a sense of satisfaction or, equally, can elicit that sinking feeling of impending frustration. Intuition is nonverbal, yet trainable. Intuition is not part of present-day machine chess”.

In the next section we present a research agenda for the data preparation required to mine vast databases of chess chunks, encoding numerous positions.

2 Distance analysis for a cognitive architecture

We are interested in the sub-conscious information processing that creates the intuition of a chess expert; a ‘locking in’ process which at start has only the
pieces present in the board and their specific squares. To conceive the board as a meaningful structure, the system has to start gathering relevant data about the deeply hidden relationships of that position.

As mentioned, chunks are stored in a long-term memory chunk network, and their information should be accessed in a fragmented manner: a bit of data triggers activation of a possible chunk, which ‘tests’ an hypothesis, by placing top-down impulsive processes looking for additional data which would be coherent with such chunk (and that would reinforce the original hypothesis).

What activates a chunk? Simon and Chase [9] and de Groot [3] had demonstrated that the saccades of players’ eyes moved mostly between related pieces, that is, pieces that are either in an attack or defense position. Let us generalize this idea by considering the concept of distance between two pieces, that is, how many movements lay from piece X to piece Y? What is the minimum number of movements for the white knight to attack the black queen? This distance metric provides direct information on the pieces that are ‘related’ (in the sense of Simon and Chase [9]): their distance is 1. Moreover, it also gives a more precise measure of the pressures involved in the chess board, by presenting those pieces that may attack (defend) each other in 2 moves, or 3 moves, or 4 moves, leading all the way to those which do not present any threat (or chance of defense) at all. So this metric captures one concrete aspect of board positions.

Figure 2 presents an example. Consider the relation between the white knight and the black queen: the first can reach (attack) the queen in 1 move, while the queen can attack it in 2 moves. This kind of relationship can be described by a graph in which each (directed weighted) edge denotes the minimum number of movements between two pieces (represented by nodes). This representational scheme seems to be in line with the view that “chess players characterize the board spatially” (de Groot [3], p.7), but it goes beyond considering the board visually as a simple Euclidean space, since the representation deals with the idiosyncrasies of each piece type’s movement capabilities.

Figure 2: White to move: mate in one. Spatial graph representations manipulated in working memory are represented by minimum distances between pieces.
Note also that many abstract relations are described in the graph. As each piece threatens to occupy some board squares in the following several moves, it is precisely which threats are perceived, which relations are ignored, and how the pieces interact as a whole that differentiates skilled from unskilled human players. A fork is an attacking relationship where a piece simultaneously threatens two opposing pieces. For instance, the graph in figure 2 displays an identified instance of a chunk of a black rook fork. But there are multiple candidate conceptions of how the pieces combine to defend and to attack the opposing camp, and this is not the only perceivable fork in the position. Chess perception is not an objective, one-to-one mapping, but a compromise, in which some features (chunks) are identified, while others are ignored. In fact, since so much structure is simply not perceived, maybe the term perception itself is somewhat misleading. In previous works, we have referred to this as multiperception, and lately to a process of pattern conception, with the term conception’s natural meaning of a conceived alternative – one specific conception – in detriment of many potential alternative conceptions [17,18]. Saariluoma [19,20] refers to roughly the same concept as apperception (see also [21,22]).

Another aspect captured by this representation is the view of analogy between chunks and analogy between chess positions. Once the distance relations are mapped, each particular piece’s identity gradually loses importance: If we look at the distance graph and discard the identity of the pieces, we can obtain a set of positions that are analogous to this one in an abstract and profound sense: positions in which the “very same” pressures emanate from a radically distinct variety of piece combinations. As a trivial example, the fork could be emanating from a black queen, and nothing essential would be changed in the graph structure – and should we imagine a whole set of transformations of this type, which change the actual board while still preserving the large scale structure of the distance graph, then after some steps we should have a board position that looks very different from the original one but which still remains in a deep sense analogous to it. This raises intriguing possibilities for new psychological experiments concerning “analogies between positions”: if chess representation in experts is encoded in a structure such as this, it can clearly be subject to empirical enquiry.

One should not prematurely presuppose that such metrics are simple to devise. There are many possible gradations of such functions, varying in terms of computation speed and precision, from an extremely fast lookup table to a full depth intractable combinatorial search. It is possible that a significant part (or even most) of the intelligence of experts comes from identifying when to select each specific gradation to use in each particular case. There is great need for this distance analysis research agenda. Let us consider 5 levels of distance metric quality:

(i) **Heuristic glance.** This level is the fastest possible distance evaluation. It can be designed by implementing a set of lookup tables with the following information: the type of piece and its square, and the destination square; the lookup table would provide a first estimate of how many movements that piece
remains from the destination. Notice the obvious limitations of this estimation, as it does not include blocking pieces, intermediate positions under immediate threat (which probably could not be used), and, most importantly, potential opponent responses. Note that this $O(1)$ computation can be performed for all piece pairs on a chess position in fractions of a second, giving an initial estimate of both attack and defense distances involved, and stressing the (direct, 1-move) immediate attacks and defenses involved.

(ii) Estimate considering blocks. Because not all 1-move attacks (or defenses) pointed out by a heuristic glance can actually be executed in a single move, the next step up would be a fast algorithm to compute the minimum path to the distance, while considering blocks, that is, occupied squares. This demands the use of a backtracking technique, and takes much more time than the heuristic glance, but could still be done in a small number of operations, since the number of backtracking operations is bounded by the number of pieces present in the board.

(iii) Estimate considering blocks and threats. This level of distance metric quality would function exactly as the ‘estimate considering blocks’, except by its consideration of potential threats by the opponent. This would require that the distance from the opponents pieces towards each position on the path be calculated, which would obviously add to processing time.

(iv) Estimate considering multiple levels of response. This estimation would be considerably more sophisticated, by employing an anytime tree searching algorithm that may consider multiple levels of opponent responses. It could take large amounts of time, but it could for some cases, and for highly salient relations, at strategic points in the game, compute the exact minimum distance between a piece and its destination.

(v) Full combinatorial search. For the sake of completeness, an extreme possibility would be the application of a full scale combinatorial search of all potential movements and all opponent responses, which would place enormous demands on computational time, but could eventually obtain winning movement combinations.

3 Psychological plausibility of the model

The cognitive chess model proposed here should operate with a mix of these levels of distance metric quality. It seems plausible that analogous mechanisms of differing levels of distance evaluation exist in the chess processing of experts. Four key reasons support this view:

First, psychological experiments have documented that chess players concentrate eye movements between related pieces; this corresponds to a heuristic glance level of distance quality, by merely perceiving that the distance between a piece and its destination equals 1. In order to perceive this, there must be eye saccades between the origin and the destination pieces. This is done instantly by human chess players, as Reingold et al. [22,23] put it, it is an “automatic and parallel” process; and this instantaneousness is also reflected on the metric.
Data Mining IV

Figure 3: Representation of a position reported by an expert chess player (After de Groot [3]). Notice that most distances are not carefully counted, while the “king trying to capture the onrushing pawn” is. This is compatible with distance analysis on multiple levels of quality.

However, on endgames with complicated positions, human chess players often reflect in a somewhat distinct manner, by carefully counting the piece distances and systematically considering potential opponent movements. A glimpse at figure 3 shows the underlying logic, where the movements of the king “trying to catch the onrushing pawn” [3] are carefully and deliberately counted, while other relations do not have such an explicit detail. At those moments in time, it seems that the subcognitive mechanisms more closely resemble those postulated as ‘estimates considering multiple levels of response’, in which distance is carefully evaluated on a square-by-square, move-by-move basis. So it may be that humans employ a mix of those levels of quality of distance evaluations. This is a second reason for distance graphs.

The third reason comes from a related feature of this representation: it enables a much more selective search than a classical tree model, because the focus here is to search movement combinations starting from piece X and leading to piece Y in order to measure the distance (obviously considering movement responses from the opponent). This is severely more restricted than iteratively searching each open movement possibility, and each open opponent response, ad nauseam.

A final reason explains what some competing theories cannot: the presence of empty spaces in chunks and in representations – as had been proposed in Charness et al. [24]. If piece A is 2 moves away from piece B in a sophisticated evaluation level, and if this is an important relation in the context of a position, then there is an empty space between A and B which is obviously relevant for representation of the ‘dynamic complex’.

4 Discussion

What kind of contribution may data mining offer to cognitive science? This paper rests on the following scientific hypothesis: Do human experts access chunks by a distance-network like model? Data mining technology may be the
key to testing this theory. But before data mining methods are applied, some data preparation is necessary, as differing distance evaluations seem to account for some notable characteristics of human play. Sometimes, when players display fast eye saccades between related pieces, as modeled by a ‘distance glance’ level of quality, they seem to occur “automatically and in parallel” [22]. At other times, it seems that players are carefully, deliberately, counting the number of movements between pieces – which corresponds to higher quality distance estimations. This type of distance evaluation also can be seen as a context-sensitive, highly selective tree-searching, a widely known fact of human play, as opposed to the exhaustive exploration of combinatorial possibilities. Finally, distances stress out some critical empty spaces in the chessboard, thereby accounting for a vital feature of human chess representations.

This hypothesis means that programs should make specific predictions about human players performance. How are specific positions interpreted by most chess experts? What are the most relevant features? Which chunks are identified? Which empty spaces should take part as vital? If in fact humans memorize boards from information related to these distance evaluations, this may clearly be tested in careful psychological experiments. For example, imagine two very distinct board positions P1 and P2 in which different sets of pieces appear. Consider the possibility that, on the surface, position P1 does not seem to have any similarity with position P2. This is simple, as P1 and P2 do not share many pieces. But, as we start to evaluate distances, both positions gradually converge to a similar abstract ‘distance graph’ between pieces. If this distance-network memory organization theory is correct, then necessarily human players must report that they see positions P1 and P2 as analogous, and maybe even occasionally report that they seem in fact to be, structurally, the same position. This conceptual model of the chess chunk can be tested empirically, by generating board positions that share a deeply similar architecture, while a shallow dissimilar layout, and experts would be able to map pieces from a board into the other. Control experiments could include positions with the same pieces of P1 or P2, but significantly distinct underlying distance structures, and subjects would not be expected to report positions as having deep similarities (despite the surface appearance brought by use of the same pieces).

There are two key contributions in this paper: first, the establishment of the deep connection between data mining methodology and the psychology of expert chess playing opens new avenues for research. And the discussion of the required data preparation methods given by the distance analysis algorithms between pieces and their conformance to the psychological literature. Humans do not play chess by searching the game tree. Humans instead seem to play chess by carefully mining the right chunks. It is thus essential to devise meticulous experiments to test this hypothesis, and this is just one kind of experiment where data mining may provide the crucial scientific evidence for the development of new theories of artificial intelligence and the cognitive sciences.
References