Research review of Deep Blue by the IBM Watson Team

In 1997 IBM's Deep Blue chess machine won a historic match against the World Chess Champion Garry Kasparov. This was a culmination in a multi-year effort to develop a machine that is capable of beating the world's best players. It all started back in the 1980s with ChipTest and Deep Thought built at Carnegie Melon University. The work was continued at the IBM T.J. Watson Research Center, where Deep Thought 2 was developed and later, the Deep Blue 1 and Deep Blue 2. The latter won the match of 1997. The win was enabled by evolution in the machine's single chip search engine and the ingenious split between hardware and software search, the use of complex evaluation function, the high level of parallelism in the system and extensive game play databases. Each of these winning implementation details are briefly summarised below.

Deep Blue contained '480 single-chip chess search engines, with 16 chess chips per processor'. Each of these chips was divided into three parts, (i) the move generator, (ii) the evaluation function and (iii) the search control. (i) The move generator 'computes all the possible moves and selects one via an arbitration network'. The computed moves are ordered as close to best-first as possible. (ii) The evaluation function 'is composed of a "fast evaluation" and a "slow evaluation"'. The fast one contains all the easily computed terms and the slow one contains terms that require iterating through each column of the board. (iii) The search control implements null-window alpha-beta search. This has reduced memory requirements since alpha and beta values are not kept in memory but it also has a number of limitations, such as needing to execute multiple searches when the exact score is needed or inability to access the transposition table. Due to these limitations, a combination of hardware and software search has been used to assess the outcomes of all the possible moves. The software search (compiled C code) was implemented as depth-limited alpha-beta search with the negamax formulation where the search depth is dependent on the credit assigned to the move which is in turn dependent on a multitude of credit generation mechanisms, such as the presence of singular moves or moves that have been enabled by previous moves, to mention just a couple.

The discussion of the credit generation mechanisms brings us to another important factor behind the success of the Deep Blue – its complex evaluation function, the function that estimates how beneficial/damaging each potential move is. Incredulously, the Deep Blue evaluation function had more than 8000 features that each move was tested against. These 'features ranged from very simple, such as a particular piece on a particular square, to very complex', such as 'rooks on files' evaluation that consisted of a multiple different sub-features whose estimations would add up to give an evaluation of the position. The evaluation function in particular has been re-assessed and tuned in each iteration of the machine to reach the winning combination in Deep Blue 2.

The fast execution of search in Deep Blue was enabled by high level of parallelism – the machine contained 30 nodes/processors with 16 chess chips each and search could be executed in multiple nodes simultaneously. There was a single master node that would control the other 29. This centralised control model had a number of drawbacks that had to be addressed. Namely, (i) controlling the distribution of work load, aka efficient load balancing; (ii) excessive load on the master node, that would create a performance bottleneck and (iii) since the information was not shared between nodes, they had to be synced up from time to time. The syncing was executed in the so called type 1 and type 2 nodes. Type 1 nodes would search through alternative moves after the first move has been examined. Type 2 nodes would search through alternatives when ‘fail high’ move has been identified. There were also type 3 nodes that would search through alternatives when all the moves ‘fail low’. The Deep Blue team measured the efficiency of the machine parallel search performance to be ‘about 8% in tactical positions and about 12% in quiet positions’. Although it could have been improved further, it was not necessary to achieve the desired outcome at the 1997 match.

Finally, in addition to the book of opening moves (about 4000 positions), Deep Blue also had access to 700,000 Grandmaster game database. During the game the board positions were tested against the entries in the databases and scored on a number of features, such as the number of times a move has been played or the level of the player(s) that used the move. It was assessed that this historic insight provided by the extended book ‘was used with good success through the matches with Kasparov’.

Overall, Deep Blue was a tremendous achievement that truly pushed the limits of what was possible at the time. It has become a significant historic milestone in the field of AI and deservedly so.