<TDNTupleAgt.docx> [FutureWork.docx](file:///C:\Users\wolfgang\AppData\Roaming\Microsoft\Word\FutureWork.docx)

TODOs for GBG-Framework – WK/03/2019

## From TR-GBG.pdf (is commented out there)

* Optional game visualization and game logging during competitions as well

## From GitHub Wiki - Research (is deleted there)

\* [Konen2018a] Konen, Wolfgang: \_General Board Game Playing as Educational Tool for AI Competition and Learning\_, Technical Report, TH Köln - University of Applied Sciences, [http://www.gm.fh-koeln.de/ciopwebpub/Kone18a.d/ToG-GBG.pdf](http://www.gm.fh-koeln.de/ciopwebpub/Kone18a.d/ToG-GBG.pdf), 2018.

## Solved items

* OK 2019-09-04: When debugging TDNTuple3Agt for game Sim, we discovered two bugs related to adaptation when a random move is encounterd:
  + trainAgent(): randLast[curPlayer] was updated only after calling adaptAgentV(), which was wrong. It leads to an adaptation when a\_t is a random move AND it leads to no adaptation in the next round when sLast[curPlayer] is the state reached by previous a\_t, but the current a\_t is greedy again. Both was wrong and both effects are now fixed by moving the assign to randLast[curPlayer] up a few lines
  + adaptAgentV(): Before, we would learn on a random move, if it resulted in a game-over state for s\_next. [The rationale behind: If we get a TicTacToe-win by a random move, we can learn from this reward, because no better outcome is possible. But even for TicTacToe this argument is flawed: If a random move would result in a tie, we would learn on this, although another better move might be available.] Anyhow, on a game like Sim, the argument is plain wrong: In Sim, if a random move results in a game-over state, this is **a loss** for the player making this move. So we should NOT learn on this, because usually a better move is possible. 🡪 Fixed by removing && !s\_next.isGameOver() from the if-clause.
  + TODO: Need to check the new behavior on all other games than Sim.
* OK Timing aspects in GBG:
  + OK Bug in QuickEval and 2048: The moves/second are 100.000 when nPly = 0, but reportedly 500.000 when nPly = 1. This cannot be, although nPly=1 uses parallel evaluation. – This can be, exactly due to the parallel evaluation on 6 cores. The time measurement is wall clock, but 6 cores are calculating moves in parallel.
  + OK Make a concept how to measure times for agent play & agent training. – Moves/second, i.e. time for agent play, is measured each time a quick evaluation is started in XArenaMenu.evaluate(). Times for training are measured in AgentBase.durationTrainingMs. Times for evaluation during training are measured in AgentBase.durationEvaluationMs.
* OK We abandon the long switch list in XArenaMenu.loadAgent by introducing a new agent member function fillParamTabsAfterLoading ()
* OK AlphaBetaAgent in C4 is now serializable, thus it can be included in tournaments:
  + We made member books transient and re-instantiated it after loading (the same way we do in constructAgent and fetchAgent 🡪 the relevant code is now concentrated in AlphaBetaAgt.instantiateAfterLoading)
  + We simplified the interface by introducing the class hierarchy   
     AlphaBetaAgent extends C4Base extends AgentBase.  
    Thus, each AlphaBetaAgent is now also an AgentBase and we can drop certain distinctions in other places: transformObjectToPlayAgent. And we can re-use many functions of AgentBase.
* OK XNTupleFuncs::makeBoardVectorEachCellDifferent() – see [here](file:///C:\user\datasets\Vorlesungen\FHK\DiplomArb\QQ2\INF-Prj-DittmarCöln2018\notes-INF-Prj-dittmar.docx#makeBoardVectorEachCellDifferent) for reasons.
* OK Make Edax in Othello fully working:
  + add methods initForNextGame() and initForNextGame(StateObservation startSO) to interface PlayAgent (see remarks in XArenaFuncs.compete(), to make the Edax-specific code lines there obsolete)
  + add method reset() to interface StateObservation() (for lastMoves, see remarks in XArenaFuncs.compete()) – or, if the problem with lastMoves is solved by the corrected lastMoves.clone() in StateObserverOthello‘s constructor, add just an assertion at the start of for-k-loop to ensure that startSO.lastMoves is always empty.
  + make Edax working for Arena.InspectGame(). This might need a new interface method PlayAgent.forceNextAction() or change the interface method getNextAction2()
  + add an assertion in Edax.getNextAction2() that the action returned is among the current available actions (a discrepancy might occur if Edax is in another state than the current state).

All the above items are now solved by the **new agent** **Edax2**, which sets the Edax board & player in each call to getNextAction2() 🡪 much simpler and no need any more for initForNextGame(), forceNextAction and StateObservationOthello.reset() (!!)

* OK Add parameter tab for Edax2
* (OK) SarsaAgt-Othello bug: When training SarsaAgt in Othello, an OutOfBound exception would occur. Reason: nTuples in NTuple2ValueFunc had length 60, but 64 was required (0,…,63 is the range of possible actions).
  + 1st fix: Fixed by letting StateObserverOthello::getAllAvailableActions() return a list of size 64 (and not 60, as before). But this is only a partial fix: If the 64 actions had numbers 10,…,73 (instead of 0,…,63), there would be still an out ofBoundException.
  + 2nd fix: Pass the list allAvailActions into the SarsaAgt constructor. SarsaAgt has private member function processAvailActions(), which finds min and max of allAvailActions and sets the Sarsa-numOutputs to max(allAvailActions)+1. This allows to pass in the available 60 Othello actions 0,…, 26,30,…,63 and still have Sarsa-numOutputs=64. But this does only work, if min(allAvailActions)≥0 and it would be a terrible waste if the int numbers of allAvailActions were 0,10,20, …, 630.
  + 3rd fix <<TODO>>: The only safe way is to use another indirection: The Sarsa-numOutputs are =allAvailActions.size and the ith output of the net stands for action allAvailActions[i]. This needs a redesign of the relevant SarsaAgt methods and of course it will invalidate all older SarsaAgt’s stored on disk 🡪 change serialVersionUID. But it is ultimately worth the effort to have a safe design which works, irrespective of the action set of ints: I. e. it is not bug-prone if a user tomorrow adds a game where the set of all available actions has the int set {-10, 0, 3, 75} 🡪 the Sarsa-numOutputs would be exactly 4.
* OK RubiksCube: TDNTuple3Agt had several bugs which hindered it from learning at all in the RubiksCube case: m\_finished not considered as break in while-loop, m\_counter not reset to 0 prior to while loop, sLast[n] not set for the very first move in a training episode 🡪 now all three items are fixed.
* OK StateObserver2048: we have member moves, but also m\_counter in ObserverBase. – We removed the obsolete ‘moves’.
* OK StateObserver2048: It is probably wrong that incrememtMoveCounter is called in **both** methods advanceDeterministic and advanceNondeterministic. – Now we call incrementMoveCounter only in method move(), which is called by both advance and advanceDeterministic.
* OK Fixed playStats.csv (class PStats): now correct for N-player games as well (N stringDescr for players, player +action added to list), start state before list. Fixed a bug w.r.t. gameScore printout. Fixed a bug w.r.t. move counter in case 2048.
* OK Removed Multi Compete from menu and from the sources (multiCompete now deprecated). It is mainly obsolete, we have the tournament system now. (And the repeated training generates for most use cases a too complicated workflow.)
* OK Removed MultiTrain from the Compete menu (we have the MultiTrain button).
* OK Fixed a bug in multiTrain: if mEvaluatorT.m\_mode == -1, an assertion would fire, because Measure oT did not receive any evaluation results. Similar for mEvaluatorQ.m\_mode==-1.
* OK Update GBG Help File to the new situation V2.0.
  + Hint to install tipps on GitHub 🡪 Help
  + Update Param tabs (with pics)
  + Add Tournament System help
* OK Evaluator2048: Check whether m\_mode is really used in eval\_agent. – m\_mode is set via Arena[Train]2048::makeEvaluator(…)
* OK Fixed a tiny bug concerning EvaluatorNim.getTooltipString()
* OK Simplify the Evaluator-classes: put more functionality and protected members centrally in abstract class Evaluator.
* OK Simplify the interface concerning the various getGameScore- and getReward-functions in StateObservation. Do we need all of them?
  + OK Classes derived from ObserverBase need only to implement so.getGameScore(StateObservation refer). The other two, so.getGameScore() and so.getGameScore(int), are implemented by ObserverBase.
  + OK so.getGameScore() is now deprecated, use so.getGameScore(so) instead.
  + [Should we abandon so.getScore(…) in favor of so.getReward(…,rgs)? – No, instead do the following:]
  + OK Classes derived from ObserverBase need not to implement any of the getReward(…) functions. ObserverBase implements default versions which just return the game score as reward. (This is for the case rgs==true. If rgs==false, a warning is issued.)
  + OK Only if the derived classes want something else to happen in case rgs==false, they need to implement getReward(StateObservation refer, boolean rgs). See StateObserver2048 for an example.
  + [Rename pa.getScore(…) to pa.getValue(…). – Not yet.]
* OK Add game-specific agents (AlphaBeta for C4, Bouton for Nim) to the agent list of certain games in order to run competitions with them (use the game-specific agents also outside QuickEval)
* OK Bug Max-N & Nim: When “use hashmap” is activated, the calculated scores are wrong. When deactivated, everything is o.k. (but it takes longer) ((5,5,5)-Nim with MAX\_MINUS=3, TreeDepth=10). If we set however TreeDepth=15, it is with “use hashmap” o.k. as well. – It is not a bug, it is a feature: For (5,5,5)-Nim the longest lasting episodes have 15 ply (each player takes 1 piece per move), so we need TreeDepth=15 for proper operation. It is just ‘luck’ that w/o “use hashmap” it works for TreeDepth=9 or 10 as well.
* OK numDescendants() has in SingleTreeNode “N += 1 + …”, but in MCTSEChanceNode “N += …”. – Clarified, SingleTreeNode was wrong. Now fixed.
* OK Distinguish somewhere in docu the different string representations an agent can have: getName, getSimpleName, stringDescr, stringDescr2, … 🡪 Table at the end of TR-GBG.pdf.
* OK Clarify whether we want to use HashMap in MaxNAgent (or only in Minimax) – We use HashMap in MaxNAgent (optionally, if set in MaxNParams), but we do not use it in ExpectimaxNAgent. We also do not use HashMap in MaxNWrapper or ExpectimaxNWrapper (HashMap could become way too big, if the wrappers are called with different StateObservation objects).
* OK Bug fix in MaxNWrapper: Since HashMap is disallowed when wrapping other agents, we set now **m\_useHashMap=false** explicitly in all MaxNWrapper constructors (regardless what is set in mPar). This leads of course to a slow-down in moves/second.
* OK Improve GUI layout for Arena and ArenaTrain
  + OK Arena: announce which column is for which player (‘Param X’ and so on from ArenaTrain is missing here).
  + OK Show the color of the player
  + OK Move button MultiTrain in the ‘Agents trained’ row.
  + OK Arena: skip the train-related things in the two rows ‘Train games’ and ‘Agents trained’
  + OK Arena: there is no way to select Human Player via agent select box. Should we allow to select all agents (or all agents which need no training) via select box?
    - OK Yes, all can be selected. The agents who need no training can directly ‘Play’ or ‘InspectV’ (they start with certain default values). The agents who need training can be selected, but they won’t run in Arena because they are in AgentState.INIT. Agents of that type can only use ‘Play’, if trained agents are loaded from disk via menu.
    - OK The Interface Arena has a method **hasTra****inRights()**, which returns false for Arena, true for ArenaTrain. If false, the GUI will skip certain elements and buttons or add certain others
  + OK: Arena: fill or delete the empty space
  + OK: Arena should have ‘Params X’ and ‘Params O’ enabled, in order to select the evaluator mode for QuickEval and to set wrapper nPly. Also, all agent parameters should be at least readable. However, only parameter for non-trainable agents (e.g. tree depth for MaxN or MCTS) are enabled, all parameter referring to training are disabled. Use [hasTrainRights()](#hasTrainRights) to decide what to enable/disable.
  + OK Arena: has now button ‘Logs’ enabled
  + OK think about nicer colors 🡪 now [THK-logo-colors](#THK_logo_colors)
* OK B**ug** **in MaxNWrapper/TDNTuple3Agt**: When the wrapped agent is TDNTuple3Agt, then it will always loose in TTT, since the returned score tuple is always (0.0,0.0). This happens only, if agent is wrapped, the unwrapped TDNTuple3Agt is OK (!). And this does not happen for TDNTuple2Agt, there we have score tuples o.k. when wrapped (??!!) – Fixed, we need in TDNTuple3Agt:: getScoreTuple(StateObservation so) the lines

sc.scTup[opponent] = m\_Net.getScoreI(bvec,opponent);  
sc.scTup[player] = -sc.scTup[opponent];

because m\_Net has – if bvec is a state where *player* has to move – only valid values if we ask it for the score from the perspective of *opponent* (this is the one who created bvec!). Still unclear why this does not pose a problem for TDNTuple2Agt, might be due to the complicated ZValue stuff there.

* OK Source folder cmaes now in library lib/cmaes.jar in order not to clutter too much the automatically generated Javadoc with cmaes-stuff. For unclear reasons it was NOT possible to package tools.cmaes.fitness (including interface IObjectiveFunction) into the same or a separate JAR. Therefore, we kept tools.cmaes.fitness, but moved the rest into JAR (If sources should be needed again, they are in lib/cmaes-sources.zip and in GitHub/GBG-backup/cmaes.)
* OK The n-tuple agent developed for C4 (Connect Four) needs to be ported to GBG.
* OK Bug fix: The current LoadAgent version ‘hangs’ if an agent with incompatible serialID is read (TTT agents TDS & TD-NTuple).
* OK Bug fix in TDAgent.java: We needed to replace the hard-coded   
  private int inpSize[] = { 6, 6, 10, 19, 13, 19, 0, 0, 0, 9 };  
  by the flexible m\_feature.getInputSize(m\_feature.getFeatmode()).
* OK TD-pars tab: Constrain ‘Feature set’ (= value for featmode) to the set of allowed featmode values for the current game (m\_feature.getAvailFeatmode()) and set the initial value to a sensible value 🡪 KG fixed this and generated a pull request
* OK: BUG1 + BUG2 fix for MCAgent: In some versions of Hex, the MC agent does not choose the best action, but **instead the worst** action (!!) This happens for (3x3, 1.), (4x4, 2.), (5,5, 1.), … and so on. The same behavior for TTT (3x3, 1.) – This was due to 2 errors: (1) in RandomSearch there could be actions on newSob even if newSob is already a game-over-state (BUG1) and (2) newSob.getGameScore has to be called with the referingState sob to get the sign right in every case (BUG2)
* OK Improved TR-GBG around game score and game value: Made it simpler!
  + Concentrate Chap. 3.3 on so.getGameScore() and so.getGameScore(sob). Explain the difference in pictures. Give examples showing explicit values for both functions.
  + Move pa.getGameScore(so) to Chap. 3.4 (Game Value), because it is a game value, not a game score (think about renaming it to pa.getGameValue).
  + Move the mind-buggling complicated so.getGameValue() and pa.estimateGameValue(so) to appendix.
* OK Some elements of measurement in methods train, multiTrain, and compete are still TicTacToe specific or at least 2-player specific. Generalize them to arbitrary 2-player games and later to arbitrary n-player games.
* OK XArenaFuncs::multiCompete writes Arena.comp.csv now to agents/<gameName> (and not to root of GBG)
* OK XArenaFuncs::multiCompete has the evaluator mode hard-wired (9). This might fail for other games where the evaluator mode 9 is not present. Extend it that it fetches static Evaluator.getDefaultEvalMode().
* OK Why has XArenaFuncs::multiCompete the evaluator part excluded in case of agent MCTS? – This was due to a wrong (incomplete) implementation of MCTSAgentT::getScore (the part with nextActionScore, same as in MCAgent, was missing). -- Additionally, there was another bug in both MCAgent and MCTSAgentT: Both would have problems when getScore(sob) was called with an already game-over sob (as it regularly happens inside Evaluator9). Then MCTSAgentT resulted in a NullpointerException, since the tree of a root state with game-over condition is never expanded in treePolicy. The fix is to check in getScore(sob) for sob.isGameOver(). If true, return directly sob.getGameScore(sob). Now both bugs are fixed and MCTS is correctly evaluated in multiCompete.
* OK Write for TR-GBG.pdf an appendix on N-tuples (with figures): Explain for what the several functions in XNTupleFuncs are needed.
* OK Bug (GUI hangs) when trying to load an old MCTS agent (v12). Happens only sometimes, only when running, not (!) in debugger. – No longer seen after transition to v13 (ParMCTS).
* OK Types.ACTIONS: replace enum (which needs to be extended whenever we need more ACTIONS) by some class construct
  + a class having an int as member
  + we delete fromString, fromVector, they are never used in GBG
  + older logs (with enum ACTIONS) are not readable anymore
* OK Bugs in Hex (for KG):
  + OK When playing a human-agent game, then the scores of the agent are not shown during play (they are however shown when replaying it with the game log)
  + OK Hex: extend to logging with subdirs 🡪 KG
  + OK Under Hex, both Arena and ArenaTrain open on “Load Agent” the wrong directory (it should be agents/Hex/<num> where <num>=BOARD\_SIZE)
* OK Saving (serializing) of things (logs, agents) was not safe, if something changes in the classes being serialized. Is it possible (with minimal effort) to read older versions of serialized objects as well? – It has become better with the proper definition of [serialVersionUID](#serialVersionUID).
  + A further improvement might be to catch an InvalidSerialVersionUID exception and react on this with transforming from an older version? – Does not work, because the class name is stored and will not fit to the new version. If we want to re-use older versions, we have to transform them, see **TransformTdAgents** for an example transforming v12-TDNTupleAgt’s to v13.
* OK Added   
   static final long **serial****VersionUID**

to serializable classes to ease the serialization process. Then the **serialVersionUID** is no longer compiler-dependent. We may add functions to a serializable class w/o unwanted change in serialVersionUID. We may add or delete single members and still read old versions, as long as the new members can be set from suitable defaults.

* OK MCTSAgent-Design-Flaw: It has params, SinglePlayer m\_mcPar and the single elements 🡪 too much!!! See if we can delete params 🡪 yes, we did. And we let the setters for single params change m\_mcPar as well. And we added getters for MCTSAgentT. Do we need setters in MCTSAgentT (or do we go always through constructor)?
  + We have some getters in MCTSAgentT, but we can use mctsPlayer.getParMCTS() for the others
  + We do not need setters.
  + We have now less parameters in MCTS since we upgraded to the ParMCTS version.
* OK MCTSParams: bug in getK\_UCT(): intValue() 🡪 now corrected to doubleValue()
* OK MCTSParams tab: when changing params and directly saving MCTS, the params are not taken. Only after a “Play” the params are taken over (!) 🡪 saveAgent in XArenaMenu now uses fetchAgents first!
* OK An object of class Evaluator is currently constructed in several places (multiCompete, multiTrain, train, and taskState TRAIN, menu item ‘Quick Evaluation’ (fct evaluate)), often with different objects (different evaluator modes), and with modes tied to game TicTacToe. Make the modes as simple as possible, then generalize it in such a way that it is for arbitrary games. Perhaps with user-adjustable mode selectors in the ‘Other pars’ tab, perhaps with sensible defaults set in the files defining constants. 🡪 see [#New Evaluator Concept](#_Evaluator_concept)
* OK During TDNTupleAgt-training: Replace the Minimax-evaluation by proper general evaluation (replace JFreeChart plot title)
* OK Disable “Output sigmoid” in TD pars when TDNTupleAgt is selected (but check it)
* OK What is NORMALIZE in TD pars? Add tooltip text. – NORMALIZE is for normalizing CurrentScore = so.getGameScore(…) in getNextAction, such that it fits to the range of the prediction of TD- or TD-NTuple-agent ([0,1] for Fermi fct in TD, [-1,1] for tanh in TDNTupleAgt)
* OK Update HelpGUI-Arena-GBG.htm
* OK [MCTSParams-vs-ParMCTS fixes](#MCTSParams_vs_ParMCTS_fixes). Now also for MCAgent (ParMC). Added member ParOther m\_oPar to MCAgent and MCTSAgentT. Now also for TDAgent: member TDParams m\_tdPar transformed to ParTD.
* OK Minimax on 4x4-Hex: InspectV shows an initial value function with some values ‘0’. Why? – Because the Minimax default depth 10 is not enough. Enhancing to depth=12 solves the problem.
* OK Extend all classes derived from XNTupleFuncs by a new member serialVersionUID. Use this serialVersionUID which is stored in the saved agents. (We need this serialVersionUID, otherwise the tiniest change in XNTupleFuncsXX will invalidate all saved agents.)
* OK When replaying a human-agent game with the game log, then any ADVANCE will move two plies forward (one agent, one human ply).[[1]](#footnote-1) When replaying an agent-agent game, it is correctly only one ply per ADVANCE. – This was a bug in the logging implementation: logManager.add(…) was only part of the if(gb.isActionReq())-part in the play-while-loop. It was missing in the human-move-part. Now solved: Added a proper statement logManager.add(…) to all GameBoard implementations in function HGameMove (or similar). This required two new getter-functions in Arena: getLogManager() and getLogSessionID(). Done.
* OK Hex:
  + OK If agents/Hex does not exist (Arena.comp.csv), then create it
  + OK Wrote in the Javadoc of StateObserverHex (a) how the tile [i,j] translates to board cells and (b) how the board numbering in XNTupleFuncs::getBoardVector translates to the board.
* OK NTuple2::getBoard is still specific to TTT (vector of length 9). Generalize! – Now set to @Deprecated and commented out, since this function is never used
* OK Replace PlayAgent::wasRandomAction() with ACTION::isRandomAction() (simpler interface, cleaner code). Delete agent’s member randomSelect, where present.
* OK Add counter ‘learning actions’ (as in [Jaskowski16], especially useful for 2048) – see TD\*.getNumLrnActions and AgentBase.getNumTrnMoves.
* OK It is very disturbing that TDAgent cannot learn the trivial 2x2 Hex well. Even with the so far best feature mode 2, the value functions for the clearly loosing start moves are still too high. Why? – And even worse, if we add with feature mode 99 a set of features exactly equivalent to TDNTupleAgt (only w/o symmetry), we do not get good results, although we have a lot of features. What is different then between TDAgent and TDNTupleAgt? – It has another sigmoid (but the difference happens also w/o sigmoid). It has another epsilon-descent. What else??  
  🡪 now largely solved, see [#Debugging TDS (TDAgent)](#_Debugging_TDS_(TDAgent)). It is the alpha-fan-in-division that matters.

### Solved items March 2019: MCTS and MCTSE

* OK EvaluatorHex uses still Minimax (and not Max-N) for certain modes. And Minimax takes forever on 5x5 Hex (Max-N probably as well) 🡪 add a warning. Now EvaluatorHex uses MaxN
* OK **Bug** **MCTS & 3x3 Hex** (2019-03-12): When MCTS iterations are set to 1000 or 10000, the first move selected by MCTS is not always a winning move (!) That is, MCTS as 1st player will often loose against Max-N, although the 1st player has a theoretic win in Hex. Win rate for 10.000: only 32% (!!) This is disturbing, because [Galitzki17, p. 31] reports a 100% win-rate for all iterations ≥ 1.200 (!!) Did we do anything harmful to MCTS after the time [Galitzki17] made his investigation?
  + If “Normalize” is activated, the low-visited nodes have always a visit count of 401. If “Normalize” is deactivated, the low-visited nodes have always a constant visit count, given a certain value of KUCT. This varies from 300-600. Why is this lower count always constant over repeats and nodes?
  + If we go to 100.000 iterations, we increase the win-rate to 100%, but why do we need so many?
  + We re-installed with “MCTS0” the MCTS-version from 2019-01-22. This version is perfect on 3x3-Hex and has not constant low-visit counts. So we compare step-by-step the differences MCTS vs. MCTS0
  + We found a slight bug in treePolicy(), but this was not the main reason
  + We found the major bug in value():  
     double v = so.getReward(so, rgs);  
    is plain wrong, it should be  
     double v = so.getReward(referingState, rgs);  
    !!!
  + Fixed also the same bug MCTSEChanceNode::value() (!)
  + Bug fixed, now the constant visit counts vanish, the InspectV values on the start board are varying as in MCTS0, but have always one of the winning cells as best action, and for MCTS against Max-N on 3x3-Hex the win-rate is again 100% (Single Compete, MCTS iter=1.000, KUCT=1.414, depth=10, with and w/o normalize).
* OK Debug MCTS for Nim: MCTS is for some Nim-configurations unreasonably bad. For example, in the game state (2,2,1) the action leading to (2,2,0) has in many situations (depending on MCTS parameter settings) NOT a high score, although it should have a high score. Clarify why this is the case and why some next actions are very seldom visited (<10 for 10000 iterations).
  + **Bug fix 2019-02-09**: One reason found. The UCT formula has to balance exploitation and exploration. This works only well (see [Browne2012, p. 4 & 7]), if q(reward), the child’s total value entering the UCT formula is mapped to [0,1]. It was before the game score ∈ [-1,1] for 2-player zero-sum games. Now we added the option to normalize (map to [0,1]) in value() and replaced the negamax-principle Δ ← –Δ by Δ ← 1–Δ. in backUp() and mctsSearch().
  + With this we get better results on the (2,2,1)-situation and a better distribution of actions over children (at least ≈100 for 10000 iterations). And also for (3,3,2) we get the winning move, if K(UCT)≤4. (But with K(UCT)=10 it gets wrong)
  + OK But still the main problem of few visits to non-max-states remains, see [notes\_MCTS.docx](file:///C:\WUTemp\FH-MassenDaten\svnSoma\trunk\doc\CaseStudies.d\201314.d\CIG2014\MCTS.literature\notes_MCTS.docx) for a deeper analysis. And both versions, MCTS with and w/o normalize are equally bad in competition: playing CompeteBoth against each other, they score 0.0, because always O wins (although X should win in perfect play) and playing against TDNT3 or Max-N[treeDepth=15] they always loose. – Now solved, with 100.000 iterations, all problems are gone. (It remains however a feature of UCT-MCTS that, if there are several equivalent winning actions, usually only one of them has a high value. With ε-greedy there is a greater chance of having high values for all winning moves.)
  + OK Switch Normalize and Selector UCT/ε-greedy/roulette-wheel added to MCTSParams. ε-greedy mechanism added, new parameter ε.
* OK Nim & MCTS: Still larger heaps have mostly wrong results, e.g. (5,5,5) or (5,5,3). Is it that the rollout is uninformative in that case? – Solved, see [here in notes\_MCTS.docx](file:///C:\WUTemp\FH-MassenDaten\svnSoma\trunk\doc\CaseStudies.d\201314.d\CIG2014\MCTS.literature\notes_MCTS.docx#MCTS_NimStartMove): 10.000 iterations are too few, with 100.000 iterations the winning move is reliably found.
* OK Add a roulette-wheel selection [Swiechowski15, Sec. 2.5.1, p. 6] to MCTS & MCTSE as an alternative to UCT selection. Is it for 2048 better to use roulette-wheel-selection instead of UCT? – Apparently no.
* OK Add all enhancements for MCTS also to MCTS-Expectimax (! Important for Daniel Weitz and EWN)
* OK 2048 & MCTSE: uses so.getGameScore() as MCTS-value. Since this is the raw score divided by MAXSCORE, it is usually a **very** small value 🡪 UCT is solely dictated by number of visits. Actually, it has with uctNormalised() for all games (!) a 2048-specific normalization to the actual high score in each move. This is however a normalization quite specific for the game 2048. Add a Normalize-SelectBox to MCTSEParams and a normalization which does for 2048 uctNormalised(), but for all other games the same normalization as in MCTS.

### Solved items December 2017

* OK Port the code from CFour-GitHub ([<Documents>/GitHub/Connect-Four/CFour](file:///C:\Users\wolfgang\Documents\GitHub\Connect-Four\CFour)) so that Connect-4 is runnable in the GBG framework
* OK Sleep duration not always right in Play Hex (e.g. two moves displayed together) – This is fixed now, it was a bug in GameBoardHex: It used JPanel::repaint() which results in a painting “sometimes later”, in this case a painting when two moves were made. Now we use JFrame::paint(), which paints at the right moment (Unfortunately, paint() causes a little flicker, but we live with this)
* Distinguish somewhere in docu the different score-getting methods 🡪 Table in the end of TR-GBG.pdf.
  + Clarify the use of pa.getScoreTuple and pa.estimateGameValueTuple 🡪 see text in Appendix B.
  + The various wrappers should use getScoreTuple and NOT estimateGameValueTuple.
  + pa.getScoreTuple means: “What is PlayAgent pa’s estimate of the final game scores?”
* OK Make a table in TR-GBG appendix listing all parallel methods.
* OK Keep Gerhard informed about GBG
* OK Make a TDNTuple2Agt [sample-TDNT2-forGitHub.agt.zip](file:///C:\Users\wolfgang\AppData\Roaming\Microsoft\Word\agents\2048\sample-TDNT2-forGitHub.agt.zip) which has all parameter settings for a ‘good’ 2048 agent, but is not yet trained and thus has a **small disk size**. Thus, we can store a sample agent on GitHub which the user can train locally.
* OK Finish MODE\_3P==2 in TDNTuple2Agt and simplify TDNTuple2Agt-code, if possible.
  + OK simpler bestScore selection (2 places) with   
    List<Types.ACTIONS> nextActions = new ArrayList<>();
  + OK replace double[] rewardArr with ScoreTuple rewardTuple.
  + OK Fixed a bug with MODE\_3P==2

Test TDNTuple2Agt again.

* OK MultiTrain-EvalMode commented out (to reduce code complexity). (We could not modify it from the GUI anyway.)
* OK Finish MCAgentN. Worked in principle for numberAgents=1, quantitative testing (should have same results as MCAgent). Bug fix for {numberAgents>1 && Wrapper nPly>0 } (ArrayOutOfBoundsException, resolved). Two parallel versions (\_PAR and \_MassivePAR), the default is \_PAR, but \_MassivePAR may be selected in source code for games like 2048.
* OK TDParams: take nPly out (default: 1), take MODE\_3P in.
* OK New Evaluator mode 11 “TDReferee.agt.zip” for TTT and Hex where an opponent is needed: Load agent “TDReferee.agt.zip” from disk (see class AgentLoader, from the game-specific dir) and play against this opponent.
* OK Quick-Eval result msg and InspectV when Wrapper nPly>0: do not print agent “MaxNWraper”, but agent “TD-Ntuple-2, nPly=3”. – OK, override getName in MaxNWrapper and ExpectimaxWrapper
* OK Bug when having AgentX=TDNT, nPly=1 and AgentO=MC, nPly=2: When Running Quick Eval for O we get a result msg TD-NTuple2[nPly=2] … Why is the wrapped agent not MC? – This was a tiny bug in wrapAgents, now fixed
* (OK) Status message when starting Quick Eval. The problem: Issuing an m\_arena.setStatusMessage(str) from XArenaMenu::evaluate() does not have immediate effect on the status bar in m\_arena. Only later, when returning, the message is displayed. Unclear why. This is in contrast with “Playing a game …” which is displayed directly. – Solved partly by issuing at least a status message on console via System.out.println().
* OK **Hierarchical param tab**: one tab for every player containing itself all agent param tabs (only in this way we ensure consistent settings, e.g. the possibility that settings in OtherPars or TDPars may be different for agent X and agent O). The steps:
  + XArenaButtons:
    - change TDParams tdPar to TDParams[] tdPar = new TDParams[numPlayer]; and similar for other params
    - in constructor, numPlayer-for-loop: tdPar[n] = new TDParams(); and similar for other params.
    - in constructor, set the lists in the param tabs (feature, eval) for every tdPar[n], oPar[n]
  + in each location where now tdPar is called (XArenaFuncs, XArenaMenu, XArenaTabs):
    - have the index n available of the agent who is meant (!)
    - change tdPar to tdPar[n]
  + XArenaTabs, constructor:
    - change the already prepared for-loop over 0 to a for-loop 0,…,numPlayer-1
    - change tdPar to tdPar[n] and similar for other params (then showParams() should work which currently would not work with other for-loop)
    - activate JTabbedPane ‘outer’
    - when Param X or other param button is pressed: select the right pane n=0,1,…,N-1 of ‘outer’
* OK Clarify: Is the parameter data flow safe, if we issue a ’play’ or ’compete’ for 2 agents of same type but with different parameters? – This is solved now by having a [hierarchical param tab](#hierarchicalParamTab): one tab for each player containing itself ALL param tabs.
* OK What is with the TD pars tab when we have a TD agent for X and a TDNTupleAgt for O? Both have their (possibly different) TD parameters (!) – This is solved now by having a [hierarchical param tab](#hierarchicalParamTab)
* OK ExpectimaxWrapper, MaxNWrapper: use wrappedAgent.estimateGameValueTuple. Check both wrappers for 2-player games
* OK Remove the use of MCAgent constructor with MCParams argument (use the one with ParMC instead)
* OK Wrapper also for InspectV
* OK Extend docu: interaction InspectV & Play, different values displayed when in InspectV or in Play mode.
* OK Nicer game board for TTT (in the spirit of Hex, merge Board and VBoard to one bigger board showing the values and the moves in color black and white).
* OK Nicer borders for the JButtons in XArenaButtons (like the Buttons in GameBoardTTT) 🡪 use SolidBorder.
* OK Nicer colors in Arena (THK-logo colors).
* OK Make a select box whether to show or not show game values during play 🡪 Show V
* OK Simplify and standardize the interface to GameBoard::updateBoard()
* OK Bug: when in InspectV, we need to hit Play 2 times to start playing. Why? – The reason was an unconditional “taskState = Task.IDLE;” added at the end of InspectGame(). Now fixed by wrapping this in an “if (taskState!=Task.PLAY)”.
* OK Bug: OtherParams, numEval seems not to be saved when saving TD-NTuple-2. Fixed: was a wrong numEval\_T in setEpiLength()

### Partly solved items

## Open items

* Java-2-Python bridge: It would be nice if we could have an agent that interfaces to Python code. See <https://pythonhosted.org/javabridge/java2python.html> for a start possibility (but check other options as well). See also [this nice set of slides](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=12&ved=2ahUKEwibnoS5mrnkAhXPalAKHXaED_0QFjALegQIABAC&url=https%3A%2F%2Fwww.eclipsecon.org%2Ffrance2015%2Fsites%2Fdefault%2Ffiles%2Fslides%2FIntegrating%2520Java%2520and%2520Python_ecf_0.pdf&usg=AOvVaw24oeA4vAsbdYy71puWgyip), local copy [here](file:///C:\Users\wolfgang\Downloads\Integrating%20Java%20and%20Python_ecf_0.pdf)). Google “java python bridge”.SortedMaxN: make a perfect agent for Sim, which scales well. While MaxN(“x”,15,true) works for K6-graph, it will get too complex with larger graphs where the depth would need further enhancement. But if we switch to a sorted representation, we could cut down hm.size dramatically.
* Need to check the new behavior of TDNTuple3Agt after [two bug fixes for adaptation after random move](#twoBugsRandomMoveAdaptation) on all other games than Sim.
* After studying the [Ludii-GUI](file:///C:\Users\wolfgang\Documents\ProjectsWK\Ludii.d), we see some options to improve our GBG-GUI (decrease the number of isolated windows, make it more easy to use):
  + Make the GameBoard window a 3-tabbed window: the 1st tab shows the game board (active during play & inspect), the 2nd tab shows the training progress (active during train & multi-train), the 3rd tab shows a scrollable message window which shows the messages during GBG’s operation (most of what goes now in the Eclipse console of the GBG program)
  + Think about integrating the Log window into the main window: The Log menu could be a sub menu of main. The Log buttons could be integrated in a row of Main GBG or in a row of the GameBoard tab in the GameBoard window.
  + Add a starter program which allows to select games, switch between different scalable versions of games (needs to make some configuration constants modifiable).
* Should we make GameBoard an abstract class (and not an interface) to bundle some functionality and members (m\_so, actionReq)? Or, alternatively, add an abstract class GameBoardBase?
* It is probably not clear which range vTable has in ACTIONS\_VT for the different agents or games: Is it [-1,1], [0,1] or something else?? Clarify the range of vTable and add suitable description in TR-GBG.pdf
* To fix the [SarsaAgt-Othello bug](#SarsaOthelloBug), we needed to let StateObserverOthello.getAllAvailableActions() return a list with size()=64 (although only 60 actions are available). Think about a safer design that the implementer cannot make this kind of mistake. Clarify which assumptions all action-consuming elements of the code have about ACTIONS: Do they expect non-negative int’s? – Partial fix only. See the [discussion here](#SarsaOthelloBug) and implement the 3rd fix described there (TODO) for a safer design.
* RubiksCube: First tests (only 10.000 training episodes) with TD-agents reveal: Both agents have a reasonable quick-eval accuracy (77% and 60% for TDNTuple2 and TDNTuple3, resp.), but if we increase ‘Wrapper nply’ to 2,3, or 4, it decreases for nPly=2 or 3. **Why?** (It goes up again for nPly=4, but then it is slow):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| nPly | 0 | 2 | 3 | 4 |
| TDNTuple2 | 77% | 37% | 35% | 89% |
| TDNTuple3 | 60% | 67% | 49% | 90% |

* It is counterintuitive and bad code design, that Evaluator.evalAgent not only returns a boolean predicate, but that it writes also on the protected members lastResult and m\_msg of Evaluator. A safer and clearer to read design would be that evalAgent returns an Object EvaluationResult, which has all the relevant data (predicate, double result, String msg) and. Such an object can be saved in mEvalResult and the getters access its members.
* Think about the else-branch (!so.isGameOver()) in TDAgent::trainAgent (!!)
* Simplify the interface PlayAgent w.r.t. getScore, getScoreTuple, estimateGameValue, estimateGameValueTuple
  + Should we rename PlayAgent::getScore to PlayAgent::getValue? – No.
  + Can we eliminate estimateGameValue() in favor of estimateGameValueTuple()?
  + Check if estimateGameValue() is correctly implemented for all agents (some, e.g. TDNTuple3Agt, use the default implementation from AgentBase).
* Check if the tree depth parameter is used consistently in all places (MaxN, ExpectimaxN, MCTS, MCTSE, wrappers)
* Add a tutorial lessson „My first GBG framework“.
* Prepare TR-GBG.pdf for publication on CIplus server.
* (OK) Student advertisement for GBG.
* TR-GBG.pdf: new sections on important reoccurring questions
  + How to find good features?
  + How to evaluate an agent?
  + OK Extend and improve Evaluator chapter in TR-GBG
  + Class diagrams
* (OK) Think about the game logic for so.getGameScore(referingState) in 3-player games 🡪 see [notes\_MCTS.docx](file:///C:\WUTemp\FH-MassenDaten\svnSoma\trunk\doc\CaseStudies.d\201314.d\CIG2014\MCTS.literature\notes_MCTS.docx), Chap. “Getting The Score For n-Player Games”. – Still some thinking about intermediate game scores to do.
* Extend the logic of getGameScore(StateObservation referingState) accordingly to 3- or 4-player games
* If MCTS has several actions with the same value (all are a ‘Win’), then it does not take the shortest path to victory. Instead it takes a random among all winning moves. 🡪 KG will think about an add-on to value function which breaks ties in favor of a shorter path.
* Hex, Multi-Competition: seems that wrong agents compete (MCTS when MC is desired) 🡪 check multiCompete, check whether its training is o.k.
* Things for MCTSExpectimax (for JK):
  + OK TreeDepth <= 3 leads to NullPointerException (s.th. around   
    return uct().treePolicy().treePolicy();) 🡪 Bug is gone in new version
  + Why is TreeNode-constructor called exactly nIterations (3500) times (at least for TreeDepth=10)?
  + Why is MCTS(Expectimax) with TreeDepth=1 not as good as MC? Which role plays K (UCT)? Does it help to set it to large values?
* ? Competitor class to make the code parts dealing with competitions more similar / more safe
* Bug (**G****UI hangs**) when loading with Hex several times a TDNTupleAgt.
  + It passes the [LoadAgent] status message, but then it returns from the action listener in such a way that the GUI does no longer respond to any events (GUI windows not visible anymore). Is it only in Hex? Although the GUI hangs, a certain program like MultiTrain comes still successfully to its end.
  + It seems to be related to a resource of Windows handles: If the computer has just been booted, it does not happen (or not often). If the program was run several times, it happens very often.
  + It seems to be related to the GameBoard realization in Hex, since it does never (or only seldom) happen in 2048 or TTT.
* Think about the parameter ‘player’ in the getGameScore and getScore functions for the different TD agents. Is it correctly and consistently used everywhere?

### Open items December 2017

* Eliminate Minimax from all code, especially evaluators, and replace it with Max-N. Later, MinimaxAgent.java can be moved to GitHub/GBG-backup.
* [GUI hangs](#GUI_hangs) more often in Hex (seldomly in other games). Clarify why!
* Open question: How do we get *in general* the game values for all the players which are not the ones to move in state sob? For 1- and 2-player games it is obvious, but for N-player games with N>2 it might be tricky. In some cases, the final game score might be a substitute. But in other games, especially those with imperfect information, it cannot be known exactly and it might require considerable calculation / modeling (with a great deal of uncertainty) to estimate it.
* (OK) Implement MCTSAgentT (with getScoreTuple) to be able to run MCTS with MaxNWrapper! – Done the cheap way by implementing MCTSAgentT::getScoreTuple() via getScore for N=1 or N=2 (but Exception is thrown for N=3 🡪 there we would need the real MCTSAgentT)
* (OK) Status message when starting Quick Eval. The problem: Issuing an m\_arena.setStatusMessage(str) from XArenaMenu::evaluate() does not have immediate effect on the status bar in m\_arena. Only later, when returning, the message is displayed. Unclear why. This is in contrast with “Playing a game …” which is displayed directly. – Solved partly by issuing at least a status message on console via System.out.println().
* Is it correct that RANDINITWEIGHTS in TDNTuple2Agt is always set to false? This means that all n-tuple weights are initialized with 0.0. We have currently no parameter in ParNT that would allow to switch RANDINITWEIGHTS on or off. If we add this, we should also add a parameter for the constant EPS=0.5 in NTuple2::initWeights(), which might be too high in many applications.

Status with 5x5 Hex and wrapAgent (nPly>0)

* When nPly>0, we currently cannot call the parallel version (evalMode=10), since the MaxN-agent would not be thread-safe. It would be thread-safe, if we delegate the wrapping to the callables (this requires nPly to be passed to the Evaluator). Think if we want to do this.

## equals() or hasEqualState()?

* Design discussion: Is it safe to have StateObservation::equals() for checking that two states are the same?
  + pro: If we have it this way, we can use all other code (e.g. Set, List) which builds upon equals()
  + con: There is always the default implementation Object::equals() which has a different meaning (‘equal’ means that two objects are at the same location in memory. Nevertheless, two different objects can have the same state). A developer of a new class implementing StateObservation might forget to re-implement equals()
* Another possibility would be to add a new function   
   hasEqualState(StateObservation arg0)

to the StateObservation interface. Then no one can forget to implement this. To have also the “pro” of the 1st solution, it is possible to define equals() as well, if it is needed:

public Boolean equals(Object arg0) {   
 return hasEqualState((StateObservation) arg0);  
}

Then two objects of class StateObservation are equal if they carry the same state, although they might be in different memory locations.

So we should change any usage of equals to the mandatory usage of hasEqualState()

Similarly with toString() and stringDescr()

## New Evaluator Concept

Status 06/2017: Where and how is Evaluator used?

* ArenaTrain, derived task TRAIN: after training, an Evaluator with mode=2 is constructed and evaluated 🡪 should be mode getQuickEvalMode()
* XArenaFuncs, train() & multiTrain():
  + m\_evaluator1: Evaluator with mode 9
  + m\_evaluator2: Evaluator with mode 2
  + m\_evaluator3: Evaluator with mode 0 (only multiTrain())
* XArenaFuncs, multiCompete():
  + m\_evaluatorX: Evaluator with default mode (for X)
  + m\_evaluatorO: E valuator with default mode (for O)
* XArenaMenu, evaluate(), which is only called by ‘Quick Evaluation’:
  + evaluator2: Evaluator with mode 2
* Disadvantage: The mode values are hard-coded (perhaps only appropriate for TicTacToe)
* The different compete menu items should be disabled for 1-player games and generalized appropriately for (n>2)- player games
* Arena2048, ArenaTrain2048: both have some strange makeEvaluator functions which compare the constant (!) ConfigEvaluator.Evaluator with 0 or 1 (🡪 compiler warning & can only be changed after recompile)
* Nicer design: it would be better to use parameter mode of Evaluator constructor / of makeEvaluator()
* EvaluatorHex: eval\_Agent chooses between three different functions competeAgainstMinimax, competeAgainstMCTS, and competeAgainstRandom at compile time
* Nicer design: Have three different Evaluator objects which decide via a mode variable which competeAgainst… function is used. This has two advantages
  + Several competeAgainst… functions can be evaluated in parallel
  + The decision, which Evaluator object should be taken can be made at run time

**The new and more general Evaluator concept**:

How should the new Evaluator usage be?

* There are five new helper functions in Evaluator interface
  + int getQuickEvalMode()
  + int getTrainEvalMode()
  + int getMultiTrainEvalMode()
  + String getPrintString() (e.g. "success rate (randomAgent, best is 0.9): ")
  + String getPlotTitle() (e.g. "success against Minimax")
* When starting ‘Quick Evaluation’, only one Evaluator with mode=getQuickEvalMode() is constructed and evaluated
* When starting TRAIN, there might be two evaluators, if getQuickEvalMode() and getTrainEvalMode() return different int’s. If they return the same int, only one evaluation is done. In any case, the QuickEval Evaluator is used to add a point to the JFreeChart plot and to check whether goal is reached (and if so, break out of training loop). The other evaluator with mode getTrainEvalMode() is only used for additional reporting at the end of training.
* When starting MULTITRAIN, there might be similarly one, two or three evaluators, depending on whether getQuickEvalMode(), getTrainEvalMode() and getMultiTrainEvalMode() return different int values.
* When starting multiCompete(), both agents, X and O, are evaluated with Evaluator having mode getQuickEvalMode().
* The helper functions getPrintString() and getPlotTitle() are for convenience, to customize printout and plot axis title.

So the simplest usage of the Evaluator concept would be: have only one mode value implemented and let getQuickEvalMode(), getTrainEvalMode() and getMultiTrainEvalMode() return this one mode value.

More advanced: two or three different evaluator modes, allowing evaluation from two or three different perspectives simultaneously.

TODO: It would be nice, if a Train Evaluator is shown in JFreeChart plot as well, but on a secondary y-axis. (This needs to understand how to plot in JFreeChart *dynamically* one item to the first, one to the secondary y-axis)

OK: Have in the “Other Pars” tab two choice boxes “Quick Eval Mode” and “Train Eval Mode”, with initial values getQuickEvalMode() and getTrainEvalMode(), resp. This offers the possibility to change these eval modes to any available eval mode (what getAvailableModes() returns). With the effect that the mode values used in ‘Quick Evaluation’ and during training change accordingly to the values in the choice boxes.[[2]](#footnote-2)

A possible later add-on: Let eval() return an object of class EvalResult which has getters getSuccess(),goalReached(), getLastResult(), getMsg(). Why? Consistent object, all information is kept together.

## MCTSParams vs. ParMCTS

Kevin Galitzki correctly pointed out that MCTSParams with its GUI being part of MCTSAgentT (and similarly in other agents) is not a good idea. Indeed, this may be the reason that GBG crashes after many MCTSAgentT constructions (“Too many handles in Windows OS”).

This is clearly a design flaw of GBG. MCTSParams is far too big to be part of other objects. And it is unpleasantly to debug, one cannot ‘see’ the parameter values in the debugger.

The new solution:

* Have for MCTSParams a sister class ParMCTS with only the parameter values as members. If there are elements static <type> DEFAULT\_... in MCTSParams, put them into ParMCTS.
* Have a constructor ParMCTS(MCTSParams) and a method MCTSParams::setFrom(ParMCTS) to copy the parameter data in either direction.
* Replace member MCTSParams in MCTSAgentT with ParMCTS. Change the serialVersionUID of MCTSAgentT to a higher ID. (Write down beforehand the parameters of all saved agents you want later to reconstruct).
* For the moment we live with the situation that older serialized objects won’t be loadable after the change XXParams 🡪 ParsXX. [[3]](#footnote-3)
* Adapt the constructors of MCTSAgentT (at least a new one with ParMCTS instead of MCTSParams)
* Eliminate all calls to MCTSParams other than the real fetching from the multi-tabbed window

Similar solutions for other parameter classes XXParams.

OK Done for MCTSAgentT, MCTSParams 🡪 ParMCTS.

OK Done for TDNTuple[2]Agt (TDParams, NTParams). In addition, with the help of TransformTdAgents and TDNTupleAgt\_v12 and a one-time call from LaunchTrainTTT constructor: all TDNTupleAgt’s saved to disk are transformed to the new version with serialVersionUID=13L. For safety, the old version v12 of TDNTupleAgt and all agent files v12 are stored in GBG-backup/v12 (can be deleted later).

OK There was a bug in the transformed TDNTupleAgt agents: The members m\_MaxGameNum, m\_NumEval and epochMax from AgentBase were not transformed properly to the new agents. Now the constructors TDNTupleAgent(TDNTupleAgent\_v12) and TDNTupleAgent\_v12(TDNTupleAgent) are extended accordingly and the whole transformation was repeated.

OK Done for MCAgent and TDAgent. In addition, TDAgent, MCAgent and MCTSAgentT get new parameter m\_oPar, for the sake of completeness

OK: MCTSExpectimaxParams and associated agents. Minimize the calls and functions using still NTParams and TDParams to only the places directly connected with the GUI. In other places we may substitute ParTD.

## Debugging TDS (TDAgent)

Kevin Galitzki found out that TDS cannot learn even the simplest Hex game, although TDNTupleAgt can. It is not well understandable that TDS has even on the most trivial 2x2 Hex such a bad behavior (value function in InspectV). This is even more pronounced when we add to FeatureHex a feature mode 99 which should be equivalent to a 4-tuple in the case 2x2. Still no good result (frankly speaking, even worse than feature mode 2).

So what are the reasons?

We compared in detail the code of TDAgent and TDNTupleAgt and found the following differences / bugs.

1. TDS had the Fermi function as **sigmoid** (if “with sigmoid”) is activated, while TDNTupleAgt has **always** tanh sigmoid. We extended TD\_func.java with a source code switch boolean FERMI\_FCT = true, false to allow either Fermi function or tanh. We corrected also a bug around (unconditional) normalization in TDAgent. Now we do a normalization only if NORMALIZE, and with min/max values appropriate to the actual FERMIC\_FCT value. With this we get for feature mode 2, 50.000 training games:
   1. good results for linear, bp, without sigmoid (value func >580 for winning tiles, <90 for loosing tiles)
   2. bad results for linear, bp, FERMI\_FCT=true sigmoid (value func >700 for winning tiles, but ≈ 600 (!) for loosing tiles)
   3. good results for linear, bp, FERMI\_FCT=false sigmoid (value func ≈ 800 for winning tiles, <90 for loosing tiles)

The results in b. are even slightly worse (≈ 900 (!) for loosing tiles) if we activate NORMALIZE.  
The results in c. are the same if we activate NORMALIZE. No wonder, since NORMALIZE does not change anything here (min and max of game score are the same as min and max of sigmoid function \in [-1,1])

So the result of this part is: Better have tanh-sigmoid (FERMI\_FCT=false) or no sigmoid, but definitely not Fermi function sigmoid

1. Rethinking of the **reward logic**: We have in both agents two places where we fetch the reward (game score from StateObservation when game is over)
   1. getNextAction
   2. trainAgent

In getNextAction, we advance the state by one of the available actions and want to select the one with highest CurrentScore. In the case of Hex (or other 2-player games) the score matters only if ‘Game over’. So we fetch in that case NewSO.getGameScore() (which is ‘-1’ in case of a win) and multiply it by (-1). Equivalently, we fetch NewSO.getGameScore(so), where ‘so’ is the predecessor of NewSO. Since ‘so’ and NewSO are always opposite, this multiplies the raw game score by (-1) as well. Either way, CurrentScore is +1 whenever a game is won.   
If the game is not over, we set CurrentScore=player\*getScore().

In trainAgent, we advance the state ‘so’ via the action suggested by getNextAction. If the game is over, we want to give a reward to the net. But the reward is not exactly the same as CurrentScore (!): reward = player\*so.getGameScore(oldSO). This is +1 for an X-win, and -1 for an O-Win (X: player=+1, O: player=-1).

Problematic in trainAgent: If game is not over, we set reward=0.0. This is no problem for 2-player games like Hex or TTT. But is it correct for 1-player games like 2048??? Shouldn’t it be the score accumulated so far? S.th. for later, we are now concerned with Hex.

Anyway, the agent is trained to let the score of predecessors of the game-over state ‘so’ become as similar as possible to the reward.

If the game is not yet over in getNextAction, we have CurrentScore=player\*getScore(). Should this be normalized as well? And if so, what is the source range, what is the target range? – TODO.

So the result of this part is: It is OK that CurrentScore and reward have slightly different formulas. We replace in all 4 instances (2 agents \* 2 functions) the complicated switch statement with one getGameScore(referingState). We have now exactly the same reward logic in TDNTupleAgt and TDAgent, if NORMALIZE==false. Needs some thinking for the case NORMALIZE==true, but for NORMALIZE==false everything is OK.

1. Why does TDS with feature mode 99 (a complete n-tuple for 2x2 board) not get the same performance as TDNTupleAgt?? Each state has a different and separate weight, so it should learn the true value function perfectly from the successor state. Why does it not work?
   1. Check the development of certain weights, whether they get a change in the same direction each time they are activated.
   2. Compare with the development of certain weights in TDNTupleAgt
   3. TDNTupleAgt has one weight table for each player, but this should be redundant here, since each board state is either only activated by X or by O.

First of all, we fixed a bug in createFeatureVector99 (wrong index calculation).

Secondly, we need much larger ALPHA, if input feature vector is long (81 values in case of feature mode 99)!! If we multiply the standard alpha values by about 100, i.e. ALPHA\_INIT=10, ALPHA\_FINAL=0.1, we get perfect results! So the **reason for not learning the trivial 2x2 board was a too small ALPHA**. This might be due to the fan-in division by length of input vector, which is active in TD\_Lin (TDAgent), but not active in TDNTupleAgt.

Thinking about the fan-in division: The divide by fan-in (which happens in TDAgent::setTDParams) is perhaps a good idea, if we expect each input unit to be (at least partly) active and if the sum of all N inputs is in the order of “N \* <AverageInput>”. But this is not the case in the “one-hot”-encoding of n-tuple-like features. There the activation sum is equal to the one active input, and alpha should NOT be divided by N. TODO: Have a parameter in the tab or a compiler switch to activate or deactivate alpha-fan-in-division.

So the result of this part is: **If we set the learning step size to ALPHA\_INIT=10, ALPHA\_FINAL=0.1**, we get perfect results for feature mode 99 on the 2x2 board. This holds for with / w/o sigmoid, linear or bp, NORMALIZE or not (!). To make adjustment of the learning rate easier in the general case, there should be a boolean switch “alpha-fan-in-division”.

Overall results: With the right ALPHA and feature mode 99 we get perfect results on 2x2 board. So it seems there is no bug in TDAgent. – But feature mode 99 is of course not transferable to larger Hex boards. Feature mode 2 shows a fair result on 2x2 board, but it is not perfect. And for larger boards we have not yet found reasonable “hand-made” features.

## Edax

<https://github.com/abulmo/edax-reversi> by Richard Delorme, 1998-2017

### Download

Go to [1 release](https://github.com/abulmo/edax-reversi/releases) (!) in GitHub (version 4.4) and download for windows edax-ms-windows.7z and eval.7z (7 MB), unpack, rename edax-4.4 to edax-4.4.exe and put data/eval.dat into the same directory.

### Workflow and associated code elements

* Edax does not have a VTable – or, more exactly, it returns a VTable with all elements 0. It returns however the next action to play.
* The interaction from Java with the executable agents/Othello/Edax/edax.exe is done in class CommandLineInteractor via members of classes java.lang.ProcessBuilder and java.lang.Process.
* Once the process is started via

process = processBuilder.start();

we can access with   
 commandLineReader = **new** CommandLineReader(process.getInputStream(), …);  
the output (!) of Edax and via

printWriter = **new** PrintWriter(process.getOutputStream(), **true**);

the input (!) of Edax. Whatever we write to printWriter will be sent to Edax. Whatever Edax outputs will be accessible from commandLineReader. commandLineReader searches the Edax output stream for occurrences of regexp pattern “[eE]dax plays …” and finds so the next move of Edax.

* When sending a command to Edax, i.e. “level = 1”, the blanks before and after “=” are important (!)
* Command “move-time 10” sets Edax to a maximum per-move time of 10 sec
* Command “setboard ............xo......ox” sets a new start square at E2-F2-E3-F3 (it complains about a ‘bad string input’, but it does the job). So the string that codes a start board is: up to 64 chars, one for each cell from A1 to H8, row-by-row, with x/o/. for player X/O/empty. Guessed from reading [the code](https://github.com/abulmo/edax-reversi/blob/master/src/board.c) in board.c.
* To enter with setboard also the player to move next: Have after ‘setboard ’ a string of exactly 65 chars. The first 64 are the pieces on the board, the 65th char is the player. Then enter ‘go’ and wait for Edax’ move.
* It is better to use in compete() ((Edax) paX).initForNewGame(startSO); and not paX = new Edax(); because the latter will result in lots of physical memory consumption if multiple games are played (!)
* The logic for deciding whether the next player has to pass is coded in StateObserverOthello.advance(): If the opponent has in the new state no possible move, then playerNextMove (what so.getNextPlayer() returns) stays at its current value.   
  IMPORTANT: In order to decide who moves next, use so.getNextPlayer() in PlayGame(), compete() or similar locations. (DON’T alternate simply between X and O.)

### Solved questions Edax

* How do we set the n-ply depth level? Via command “level 1”? – Yes, also “level 0” is possible. The default seems to be 21.
* Wouldn’t it result in a simpler interface, if we replace all the logic in Edax.getNextAction2() by

lastEdaxMove = commandLineInteractor.sendAndAwait("go");

We redesigned with **Edax2** a new Edax agent that has a much simpler interface, since it always sets the Edax board according to the actual state & player and then issues a “go”.

* For compete() we need the possibility to initialize Edax for a new game with a start state different from the default start state, i.e.

**if** (paX **instanceof** Edax) ((Edax) paX).initForNewGame(startSO);

With the new Edax2 this is no longer necessary: we set the Edax board within every getNextAction2(). No special logic is required for compete, inspect and even not for starting play out of inspect (which was impossible for old Edax)

* Edax2 is now serializable, so it can be part of Tournaments (Edax cannot be saved)
* It would be nicer, if level and move-time were adjustable from within GBG 🡪 Edax2, utilizes the Edax param tab.

### Open questions Edax

* What is the logic / the design principle behind TimeOut?? What should GBG do if Edax returns with a TimeOut-Exception?
* When playing Edax vs Edax, a TimeOut will regularly occur. We can avoid if, if we send command “move-time 10” to Edax in the Edax constructor.
* Why should we need then TimeOut at all??
* There occur still from time to time TimeOuts (Edax & Edax2)

## Tournament System

### Workflow and associated code elements

XArenaMenu “Start Tournament System” 🡪

constructor TSSettingsGUI2 🡪 constructor TSAgentManager 🡪 various TSAgentManager.addAgent, but with PlayAgent = null

finally the GUI TSSettingsGUI2 comes up and agents for tournament may be selected.

If the button “Load Agent from Disk” is pressed, then loadAgentsFromDisk() 🡪 mArena.tdAgentIO.loadMultipleGBGAgent() is called. This adds one or multiple agent names “HDD <filename>” to the component, which can be selected or deselected.

How are the agents added? – They are returned in a TSDiskAgentDataTransfer object which has members String[] filenames and PlayAgent[] agents. These arrays are filled in loadMultipleGBGAgent via TSDiskAgentDataTransfer.addAgent(filename,pa).

Once the “Start” button is pressed:

TSSettingsGUI2.**playPressed**(), normally with doubleRoundRobin selected.

The number of games (episodes) per match is set in results.numberOfGames via mTSAgentManager.setNumberOfGames().  
Some diagnostic info is printed, then mTSAgentManager is handed over to mArena.tournamentAgentManager and mArena.taskState is set to TRNEMNT

Arena.run(), switch **TRNEMNT** 🡪

lockToCompete() 🡪 The storage for all timers is reserved:   
 results.timeStorage = new TSTimeStorage[results.gamePlan.length][numPlayers];  
where gamePlan.length is M\*(M-1) if M = #agents. And numPlayers is 1 or 2. If numberOfEpisodes>1, then all episodes for a specific match will accumulate their timings in the same timeStorage array element.

while-loop (tournamentAgentManager.hasNextGame()) 🡪

tournamentAgentManager.getNextCompetitionTeam (still with mPlayAgent=null for the standard agents)

NOTE: if results.numberOfGames >1, e.g. 3, then hasNextGame() will return each individual agent pair “A vs. B” 3 times in a row (!)   
XArenaFuncs.**singleCompeteBaseTS** 🡪 the strings for AgentX and AgentO are inferred from dataTS.nextTeam[i].getAgentType().

How are agents set? – If the agents are HDD agents, the PlayAgents are taken from dataTS…getPlayAgent(). If the agents are standard agents, the PlayAgents are fetched via fetchAgents(). For this to work, the method xab.enableTournamentRemoteData is called beforehand, which sets String[] selectedAgentTypes and selectedAgentTypes [i] is the string returned by getSelectedAgent(i).

end while-loop

When the tournament (the while-loop) is done, Arena.run() calls TSAgentManager.**makeStats()**: This is quite a long method which is responsible for analyzing the tournament results and displaying them in a **TSResultWindow**.

makeStats() calculates in rowData1 and rowData3 the values for WTL Table and Score Table.

makeStats() calculates in sections “detailed time table” and “simple time table” the data for both forms of the Time Table: This is done by aggregating data from [results.timeStorage](#resultsTimeStorage)[i][j]. The different timeHelper objects are elements from array simpleTimes[] which has [M](#M_numAgents) elements. They are used for aggregating the timings for each agent.

Keep in mind: makeStats() is called each time the Start button is hit and each time a tournament file is loaded from disk. This will generate each time a new (!) TSResultWindow, but certain global variables in TSAgentManager, e.g. results, exist only once and will be overwritten. Is this safe in every aspect? Is the TSResultWindow in all aspects truly independent from the content of results or other variables in makeStats()?

Understanding the columns in Time Table:

|  |  |  |  |
| --- | --- | --- | --- |
| Column | method in TSTimeStorage | timer list in TSTimeStorage |  |
| Fastest Move | getMinTimeForGameMS | measuredTimesInNS, min |  |
| Slowest Move | getMaxTimeForGameMS | measuredTimesInNS, max |  |
| Average Move | getAverageTimeForGameMS | measuredTimesInNS, average |  |
| Median Move | getMedianTimeForGameMS | measuredTimesInNS, median |  |
| Average Episode | getAverageRoundTimeMS | episodeTimesinNS, average |  |
| Median Episode | getMedianRoundTimeMS | episodeTimesinNS, median |  |
| Total Time | getTotalTimeMS | measuredTimesInNS, Σ |  |

Each time an agent makes a move, the dynamic list measuredTimesInNS in the relevant TSTimeStorage object gets a new element containing the nanoseconds needed for that move. Each time an episode finishes, the dynamic list episodeTimesInNS in the relevant TSTimeStorage object gets a new element containing the nanoseconds needed for that episode.

Note: The average and median values per agent are calculated in makeStats() in a two-stage process: First, in section “detailed time table”, the average / median over episodes is taken and secondly, in section “simple time table”, the average over matches (that the agent was part of) is taken.   
I.e.: Average Move = 〈〈move time〉episodes〉matches. This is exactly the average over all episodes in all matches, if the number of episodes per matches is constant for all matches, which is the case here. The median is calculated the same way, which is NOT exactly the same, because the first median selects one value for each match and it is only this reduced set for which the second median selects a value. The calculated median will be thus a bit coarser than the exact median. But the difference is probably not too big.

Let E = ”Episodes per Match” and (M-1) = “number of **other** agents”. Then the following facts about the Time Tables hold:

* Simple Time Table: 2\*(M-1)\*E \* ”Average Episode” = “Total Time”
* Detailed Time Table: E \* ”Average Episode” = “Total Time”

(The first line is because each agent plays two times (one as 1st, one as 2nd) against each other agent and not against itself)

### Bug Expectimax-N in Tournament Nim

Description: When a Nim-tournament is started with standard agents Random and Max-N selected, then the tournament will construct instead of Max-N an Expectimax-N agent and run (repeatedly, another bug) into an exception.

Solution: when we arrive in the while-loop the string AgentO is “Expectimax-N” (and not “Max-N” as it should). The reason for this wrong behavior is unsafe code in TSSettingsGUI2 constructor

mTSAgentManager.addAgent("Standard MaxN", Types.GUI\_AGENT\_LIST[2], …

Since GUI\_AGENT\_LIST may change (here we commented “Minimax” out), the name stored in GUI\_AGENT\_LIST[2] may be wrong. It is better to use the explicit agent name

mTSAgentManager.addAgent("Standard MaxN", “Max-N”, …

here and in similar places.

To fix the 2nd part of the bug (singleCompeteBaseTS would repeatedly fire the Expectimax-RuntimeException and display it in a MessageBox, but would not terminate the tournament), we let the Arena.run-part, switch TRNEMNT, listen to what singleCompeteBaseTS returns: if it is the code 43 (what the MessageBox-catch-part returns), then we break out of while-loop and thus finish the tournament.

### Bug TDNTuple3Agt not loadable in Tournament

Description: When a Nim-tournament is started and one tries to load a TDNTuple3Agt from disk an Exception occurs.

Solution: When pressing “Load Agents from Disk”, then LoadSaveGBG.loadMultipleGBGAgent() is called. This has the section within the try-catch-block where the Object obj read from ObjectInputStream ois is transformed to a valid PlayAgent pa. This section was outdated since it had not yet TDNTuple3Agt and SarsaAgt included. To maintain the relevant code only once, we moved the similar try-catch-section from loadGBGAgent (which was up-to-date) into a new private method **transformObjectToPlayAgent()** and call this method from both loading methods.

A 2nd bug was that an unsuccessful agent loading resulted after the loading exception in a later crash of the tournament system (null pointer exception). We fixed this by letting transformObjectToPlayAgent() re-throw the displayed exception, which is then caught by TSSettingsGUI2.loadAgentsFromDisk() and leads to an early return with a “No agent loaded from disk" warning printed to System.out.

### Other Items

* It should be kept in mind that standard agents are fetched via XArenaFuncs.fetchAgents. This means that the parameters are fetched from the param tabs. They are only “standard” if nothing in the param tabs was changed by the user.
* OK The GUIs for TSSettingsGUI2 and for TSResultWindow (“Tournament Results”) are always too big, because the size of the Windows task bar is not subtracted. – Fixed for TSSettingsGUI2 by adding a method adjustComponentHeight() which sets the component height at the end of constructor via setSize and setBounds to not more than 95% screen height (we do NOT want to let the window become bigger than this, instead we use the scrollbar). This method adjustComponentHeight() is also called after various pack() statements. The same for TSResultWindow, here we need also to adjust the width.
* OK It would be nice if the text “Show …” / “Hide …” would change depending on context.
* OK Rename “Games per Match” to “Episodes per Match” (and similarly the relevant names in the source, e.g. numberOfGames)
* Add an option to TSSettingsGUI2, that the game plan is such that each agent plays against itself as well (the diagonal). In this way we could routinely measure, if for identical agents either the first or the second player has an advantage.
* Add tooltip texts where needed.
* OK Can we set **tooltips for the table heads**? – Yes, we found a way via <https://docs.oracle.com/javase/tutorial/uiswing/components/table.html#headertooltip> and implemented it for Time Table (look for columnTimeToolTips in TSResultWindow). Be sure that columnTimeToolTips has as many entries as the table has columns, otherwise lots of RuntimeExceptions may fire (!)
* Be aware that we now changed the method $$$setupUI$$$() at various places.   
   if (showjspSCR) showHideTableSCRButton.setText("Hide Score Table");

else showHideTableSCRButton.setText("Show Score Table");  
and similar. And

tableTimeDetail = new JTable(){

//Implement table header tool tips. …  
So it should not be generated again by IntelliJ IDEA GUI Designer (or you have to backup the hand-made changes before and restore them later)

* OK The normal return value of singleCompeteBaseTS is questionable (it tests a double ==1.0 or ==-1.0). – Well, it was not wrong in the special case competeNum=1 and the games reward being exactly 1.0 or -1.0. But we changed it now to work more generally for all values of competeNum and arbitrary rewards: We form the difference between X-winrate and O-winrate and test this on being <0, ==0 or >0.
* OK XArenaFuncs.compete and competeTS seem not very different. Compare carefully and see if we can integrate the new functionality of compteTS (mainly the TSTimeStorage stuff) can be added to compete, so that we have only one method to store and to maintain. – Done, competeTS is now commented out.
* OK Is it sensible that even for Games per Match > 1 the function compete is called with competeNum=1? – Yes, it is, if we want consistent time measurements and control over start states from the tournament side.

Related to TSResultWindow:

* OK What exactly is in the Score Table? – The WTL Score for each match.
* OK What exactly is the WTL Score? – The **WTL Score for a match** is   
   +1.0\*(#wins)+0.5\*(#ties)+0.0\*(#losses)   
  with wins/losses from the perspective of the first agent (row agent). Calculated in rowData3 in makeStats().  
  The **WTL Score for an agent** (shown in the Ranking Table) is the same, but aggregated over all matches that this agent has played in the tournament. If we divide this agent WTL Score by #games that this agent played, we get the **WonGameRatio** shown in the Ranking Table as well.
* OK Show Time Table: Why have some agents Total Time =0.0 although average episode time >0.0? – This was a bug: getTotalTimeMS used to return a long, which is wrong. Now it returns a double and Total Time is no longer 0.0.
* OK Show Time Table: Why is Average Episode and Median Episode very often exactly the same? – This was a bug: in makeStats, section “simplified time data” it was always calculateMedian(…). Now changed 2x to calculateAverage(…), and added this new method. Both calculate… methods moved to tools.Utils.
* OK Add an option to the scatterplot to toggle logarithmic scaling for time axis on/off. 🡪 Checkbox “log X”.
* Better understand the heat maps.
* How exactly are the random start moves selected?
* The [makeStats()-issue](#makeStatsIssue)

1. And when logging a 1-player human game there was no action recorded at all (!). [↑](#footnote-ref-1)
2. The third eval mode getMultiTrainEvalMode() is without choice box because it is probably only seldom used. [↑](#footnote-ref-2)
3. Several ideas to recover older saved agent versions do not work:

   * **Idea 1**: Save beforehand a sample MCTSAgentT with old serialVersionUID. Try to write code to load such an older version:
     + Catch the InvalidVersionUID exception
     + If we are in the catch block, try to get the version ID of the older object, cast it to a loader class MCTSAgentT\_v12 (a copy of the old class with old serialVersionID and old MCTSParams)
     + Copy from MCTSAgentT\_v12 to MCTSAgentT (is a shallow copy enough for all parts that did not change?). Use constructor ParMCTS(MCTSParams) for the part that changed.
   * A first analysis shows that it does not work this way:
     + The InvalidClassException is already thrown at ois.readObject() when the serialVersionUID of the class on file and the serialVersionUID of the local class MCTSAgentT do not match. That is, as soon as we change this, the object on disk is no longer readable.
     + The serialization mechanism can handle adding or deletion of class members, but not a change of field type. **Idea2**: But wait: If serialization can read an older MCTSAgentT object (with MCTSParams object but no ParMCTS object) when the serialVersionUID does not change (and it initializes ParMCTS with null), then we can read an older version w/o change in serialVersionUID. – No, this does not work really, since if the new local class has no MCTSParams object, then the information on the old MCTSAgentT object on disk w.r.t. MCTSParams is ignored (not read in), and thus no transformation is possible.
     + The only way would be to override the readObject function for the agents.   
       Or to transform any agent object to a special load/save object, e.g. MCTSAgentT\_v12, MCTSAgentT\_v13 prior to saving it to disk. If objects are saved to disk in this way, we first load the object from disk (deserialization), then check in a long switch statement the class of the loaded object (instanceof) and transform according to this class on to the ‘real’ class MCTSAgentT by copying (cannot be done through a simple cast).
     + This would require a lot of work and some maintenance costs. Perhaps something for later.

   [↑](#footnote-ref-3)