Review Feedback and Response

Paper ID	306
Paper authors	Dhirendra Singh, Sebastian Sardina, Lin Padgham, Geoff James
Paper title	A BDI Learning Agent for Environments with Changing Dynamics
Paper subtitle	
How are the claims of the paper evaluated?	Empirically (e.g., experiments)
Paper status	First Ballot
Keywords	Agent theories, Models and Architectures::BDI ** Learning and Adaptation::Single agent Learning
Abstract	We propose enhancements to a framework that integrates learning capabilities to improve plan selection in the successful and popular Belief-Desire-Intentions agent programming paradigm. In learning which plan to select, a crucial issue in the online setting is how much to trust what has been learnt so far (and therefore exploit it) versus how much to explore to further improve the learning. In this paper we construct a confidence measure based on a previously used notion of stability in the outcomes observed for a particular plan, combined with a consideration of the extent to which new worlds are being witnessed by the plan. This new measure dynamically adjusts based on agent performance, allowing in principle, infinitely many learning phases. Additionally, it scales up irrespective of the complexity of the goal-plan hierarchy implicit in the agent\'s plan library. We demonstrate the utility of our approach with results obtained in a practical energy storage domain.
Download	

Comments to author(s)
SUMMARY:
-Describe the paper in 2-3 sentences

The paper describes an approach for enhancing the BDI model with learning capabilities. Concretely, the proposed mechanism aims at supporting better plan choices for a given goal by learning in what world states which plans have been most successful.

RELEVANCE:
- Is the work relevant to AAMAS 2011? For example, does the paper describe:
* an implemented agent system
* theoretical work relevant to autonomous agents
* theoretical or applied work relevant to multiagent systems
* methodologies or languages that can be used to construct such systems

The work is relevant to AAMAS 2011. It consists of a conceptual framework as well as an implemented system that demonstrates the capabilities of the approach.

ORIGINALITY:
- Does the paper clearly point out differences from related research?
- Are the problems or approaches new? For example, does the paper:
* address a new problem or one that has not been studied in much depth?
* introduce an interesting research paradigm?
* describe an innovative combination of techniques from different disciplines?
* introduce an area that appears promising, or might stimulate others to develop promising alternatives?

The paper builds heavily on previous work of the authors and is thus mainly incremental. The main difference to earlier publications is that the learining approach now is capable of dealing with changing environments. Actually, I was a bit astonished to read this, because without being able to react on environmental dynamics the approach is quite limited.

SIGNIFICANCE
- Is the work important?
- Does it advance the state of the art?
- Does the paper stimulate discussion of important issues or alternative points of view?

The general idea of combining BDI with learning facilities is very good and important. It would be very useful for many domains if plan decisions could be quided be past experiences. The practical approach of the paper shows that this combination is possible and also demonstrates how it works using a small example application, but it does not discuss much the applicability of the approach. In which domains would it be beneficial using the approach? What are the inherent limitations of the approach e.g. with respect to the size of problem?

In the related work part you should discuss the relationship with [1].

[1] Folk Psychology for Human Modelling: Extending the BDI Paradigm
Emma Norling AAMAS 2004

TECHNICAL QUALITY
- Is the paper technically sound, with compelling arguments?
- Is there a careful evaluation of the proposed method and the results?
- Does the paper carefully evaluate the strengths and limitations of its contributions?
- Does the paper offer a new form of evidence in support of or against a

well known technique?
- Does the paper add to our understanding of some aspect of agent systems?
- [Especially for methodologies/languages] is there a clear linkage [conceptual, comparative, evaluative] to current practice?
- If the paper describes an application, is there:
* a clear and compelling motivation for why an autonomous agent is necessary?
* a clear motivation for why a multiagent approach is appropriate?
* a careful description of the design and implementation of the system?
* a thorough evaluation of the system with respect to a clearly-stated set of functional and quality requirements?

The technical quality of the paper is good and it presents the approach in a concise and understandable way. Despite this fact, I have several questions that have not been discussed thoroughly in the paper.
- The definition of plan stability says that a plan is considered as stable in a world state w if the success rate of the plan is changing below a threshold. According to what factors this threshold is chosen and how much does the choice of this threshold influence the performance of the learning algorithm?
- At what time points is stability recorded? Every time a plan fails or after a goal fails?
- A very important point is how a world state is defined, but no definition or discussion is given with respect to this issue. Does the world state encompass the whole knowledge of the agent or does it only include the parameters of a goal or plan? If the whole knowlegde of the agent is used the state explosion will lead to very inefficient learning due to the fact that always new world states witnessed. On the other hand, if only plan/goal parameters are used the questions arises if the other knowledge can be considered as obsolete for learning. Maybe there should be a way to define exactly which parts of the agent's knowledge are part of the world state. I also think that it could be an idea to think about ways of generalizing world states to allow the agent to act also in unknown world states good by knowing what it did in the most similar world state.
- The the energy storage application aims at finding a good configuration of the modules for a given rate. This sounds as if a traditional planning mechanism would be suited very well for solving the problem. So why using a learning BDI controller is better than using a planner?

QUALITY OF PRESENTATION
- Is the paper clearly written?
- Does the paper motivate the research?
- Are results clearly described and evaluated?
- Is the paper well organized?

The paper is clearly written and reads very good. The results are precisely described and also evaluated by a case study. A discussion of limitations is given but without dealing with topic how efficient the approach is. How much time of the normal execution has to be spent for learning, i.e. does learning heavily impacts the reactivity of the agent? How large problems can be that can be dealt with? Is it possible to use the approach selectively for those parts where learning is beneficial and use standard BDI for the rest of the agent?

Summary of review
The paper presents an interesting approach for integrating learning with BDI, The approach is incremental and does not discuss several important aspects of the proposed solution.

"Author Response (optional)
Authors response by Dhirendra Singh (2010-12-02 03:23:43) to review
Review Evaluation: Very helpful

Lower stability threshold implies greater accuracy (more samples) and vice-versa; while small variations do not seem to impact learning, a deeper analysis is required. Stability is assessed every time a leaf-plan terminates a trace.

The world state includes parameters relevant for the goal and plan (including descendants). This can be programmer specified, or calculated from variables accessed.

Selective learning is a great suggestion (thanks!); while we have not done that yet, it is definitively possible.

Learning does not impact reactivity; the main overhead is in rebuilding decision trees: here an incremental approach would help.

Automated planning requires a (non-changing) model of the world's dynamics, which we do not have here.

Regarding environment dynamics: difference from traditional learning is that the learning task is not stationary.

Comments to author(s)
SUMMARY:

This paper is presents a learning framework for improving the plan selection process of a BDI agent. The authors propose a new confidence measure that allows learning in domains with changing dynamics. The use of the framework is demonstrated in an energy storage domain.

RELEVANCE:
- Is the work relevant to AAMAS 2011? For example, does the paper describe:
* an implemented agent system
* theoretical work relevant to autonomous agents
* theoretical or applied work relevant to multiagent systems
* methodologies or languages that can be used to construct such systems

This work is relevant to AAMAS as it provides theoretical work in individual agent reasoning using the BDI paradigm, coupled with a practical implementation of the proposed system.

ORIGINALITY:
- Does the paper clearly point out differences from related research?
- Are the problems or approaches new? For example, does the paper:
* address a new problem or one that has not been studied in much depth?
* introduce an interesting research paradigm?
* describe an innovative combination of techniques from different disciplines?
* introduce an area that appears promising, or might stimulate others to develop promising alternatives?

This work is original. The authors provide justification based on the restrictions of previous approaches, and they clearly describe their main contributions. In addition to this, related work is discussed in detail, explaining how this approach relates to other work in the literature.

SIGNIFICANCE
- Is the work important?
- Does it advance the state of the art?
- Does the paper stimulate discussion of important issues or alternative points of view?

This work advances the state of the art in one aspect of "BDI learning". The authors provide compelling arguments regarding how the proposed confidence measure overcomes limitations of previous work. The presented experimentation demonstrates the performance of the proposed system in specific cases (which demonstrate how the system overcomes these limitations). However, the paper lacks a comparative evaluation against other common learning algorithms from the literature (beyond the "BDI learning" approach).

TECHNICAL QUALITY
- Is the paper technically sound, with compelling arguments?
- Is there a careful evaluation of the proposed method and the results?
- Does the paper carefully evaluate the strengths and limitations of its contributions?
- Does the paper offer a new form of evidence in support of or against a well known technique?
- Does the paper add to our understanding of some aspect of agent systems?
- [Especially for methodologies/languages] is there a clear linkage [conceptual, comparative, evaluative] to current practice?
- If the paper describes an application, is there:
* a clear and compelling motivation for why an autonomous agent is necessary?
* a clear motivation for why a multiagent approach is appropriate?
* a careful description of the design and implementation of the system?
* a thorough evaluation of the system with respect to a clearly-stated set of functional and quality requirements?

The authors justify the proposed dynamic confidence measure by describing the limitations of the previous measures in the literature. The technical parts of the paper all seem correct. There is also a demonstration of the behaviour of the system in specific cases (with respect to the identified flaws in the literature), and extensive discussion explaining the behaviour of the system. The discussion provides compelling arguments in favour of the approach (but no formal analytical results).

QUALITY OF PRESENTATION
- Is the paper clearly written?
- Does the paper motivate the research?
- Are results clearly described and evaluated?
- Is the paper well organized?

The paper is clearly written and well organised. The authors clearly describe the limitations of previous work in BDI learning, their contributions and their approach. The experimentation set-up is also clearly stated. The experimental results describe the average performance of the system based on 5 runs for each experiment. Additional measures regarding the reliability of the results, e.g. statistical variance, should be provided to indicate the significance of the results.

Specific Comments:

The paper is lacking a comparative empirical investigation of the efficiency of the proposed method (rather than a demonstration of its behaviour) against learning methods outside the "BDI learning" literature. The performance of the method could be compared to any general learning method from the literature (learning which (applicable) plan should be selected in a state). An example would be a reinforcement learning algorithm, based on a reward that is calculated depending on the success of the execution trace (assigned to individual decisions depending on which plans are responsible for the failure).

There should also be some discussion on how the plan selection process fits into the BDI architecture (e.g. when an execution trace fails, dropping goals, etc), especially regarding the example domain, as the presentation of the energy storage application seems to focus entirely on plan selection, without some intuition on how this process fits into the BDI reasoning scheme.

The experimental results for 5 runs for each experiment provide an indication of the behaviour of the system, but it is unclear how significant and reliable these results are.

Intuition on the importance and sensitivity of the parameters, e.g. \alpha, and how changing their values affects the learning process would be helpful.

The total number of states (13.7 million states) seems quite high. It would be a good intuition if the number of distinct states that were observed during the experimentation was provided as well (at least as a measure of a rough number of states which can treated by the system).

A formal definition of 'stable(P,w)' and 'mathcal{P}(P,w)' should be included (both in section 3).

Minor Comments:

The notions 'world' and 'state' should not be used interchangeably

"(i.e., ...)" to (i.e. ...), same for "e.g.,"

Section 1:

- "operate less well"

- "vs" to "vs." or versus

Section 2:

- "to improve its performance over time '\n',"

- "Note that the classical boolean context ... plans in various world states." commas needed, especially " initial, necessary, but not sufficient"

- "Since the decision tree inductive bias is" "bias imposes"?

- Regarding execution trace \lambda: is w' = w_2?

- What does the grey highlighting present in Figure 1? Is it the trace \lambda from the end of Section 2? This should be clarified.

Section 3:

- "is appended against it" to "to it"

- "plan would be selected in all worlds in which it is reachable" applicable?

- "if the renewable generation is high relative to the building loads" "high, relative"?

Section 5:

- Footnote 3: "Set*" is not defined

- Intervals '[0 : 3]' to '[0, 3]'

- Reference [17], capitalise MDP and BDI

Summary of review
SUMMARY OF REVIEW:

This paper presents a learning framework for improving the plan selection process of a BDI agent. The main contribution of this work is the new confidence measure that allows for learning in domains with changing dynamics. The authors describe how their approach overcomes limitations of previous work and demonstrates its behaviour in different relevant cases. The paper is well-presented and the contributions are clearly stated. However, the significance of the work is limited by the lack of a comparative evaluation against standard learning methods from the literature adapted to the problem of plan selection by a BDI agent.

"Author Response (optional)
Authors response by Dhirendra Singh (2010-12-02 03:26:08) to review
Review Evaluation: Very helpful

Good point: indeed our hierarchical learning shares similarities with Diettrich's MAXQ approach that forseeably could replace the decision-tree framework. Note, however, our focus has been the principled integration of learning with BDI, and less on the learning technique.

We shall explain the "plan selection process" clearer as well as discuss impact of parameters' values on learning: for stability threshold, see answer to Review1; regarding \alpha, it is generally not critical to performance if there are many decision paths in the hierarchy and many states in which plans apply.

The BDI execution process is unchanged: standard plan selection is simply replaced by the decision tree and probabilistic selection.

Regarding states: programmed context conditions constrain 13.7M states to roughly 1.5M.

Comments to author(s)
This paper specifies and deploys a measure which aids agents in learning how much confidence to place in alternative plans for achieving their goals. It takes into account the context in which the experience was learnt, while also allowing context-independent information to be built up. It extends existing work on choosing plans based on decision trees.

In general, I found this to be a well written, interesting and convincing paper. There are weaknesses, as a couple of important technical points are not explained clearly enough, the motivating context is not sufficiently explicated, and I dispute use of particular terms, but overall the approach seems effective and practical for applications with certain characteristics. Specific details follow.

The technical point I found most unclear, and affected my understanding of the approach, was "stability". At the start of Section 3, a definition which largely makes sense but implies that only failed plans can be stable ("A failed plan P is considered to be stable..."). Later in the section, the authors say "When a plan finally succeeds, we take an optimistic view and record 1 (i.e., full stability) against it." This seems contradictory and I do not understand. By the definition and the equation below it, "1" appears to imply that all applicable plans consistently failed. Moreover, the authors say "Leaf plan nodes... make no choices so their degree is simply 1." But, by the equation, wouldn't it be 1 only if the leaf plan was stable, and 0 otherwise? Finally, the definition asserts stability where "the rate of success of P in w is changing below a certain threshold" - but what about if it is constant and below that threshold? Do you mean "remains" rather than "is changing"? In general, this concept needs clarifying.

I found aspects of the case study unclear. In particular, do the high-level plans which achieve a goal G(r,5,s) configure all batteries in the system, or just one battery? If the former, what form does the plan take? If the latter, why would learning about configuring one battery be different about learning about another? Also, what are the "possible requests" referred to in Experiment Setup, and what is the "overall battery response"?

Starting from the description of the augmented BDI system in Section 2, I was unclear what happened to the context conditions originally designed for plans (before learning). At one point, the authors refer to the "plan's (real) context condition" and I guess this is what is being referred to (please clarify), but I am unsure whether the context condition is still checked before/after the decision tree or confidence measure is applied?

It would help to have a stronger motivation and clearer setting for the work, especially in the introduction. Section 2 goes some way to clarifying where the approach would be useful, and a little more can be inferred from the case study, but the authors should make clear up front what the set-up is. In particular, they seem to assume an agent with multiple plans available for each goal, and that those plans do not just divide up the space of possible contexts but instead there are multiple plans available in any given context (for a given goal). This does not seem intuitively the way an agent would be designed in many (most?) cases, so the authors should explain what the set-up is (i.e. is it as I just described?) and why it is a realistic set-up in some applications (i.e. applications with what characteristics?)

Section 3 discusses world-based confidence measure which (despite its name) measures confidence independently from the world in which success is judged. Are all plans equally world-dependent or world-independent? I could envisage plans whose success depends critically on current conditions and others where the general method is successful relatively independent of context, but the equation/mechanism does not seem to account for any variation. If there is a simple answer to this, it would be good to clarify.

In Algorithm 1, why is d not a parameter of RecordDegreeStabilityInTrace (as it is in RecordDegreeStability)? What is the meaning of it being omitted?

In Sections 2 and 3, there is a mixture of notations between w', w", ... and w_1, w_2, ... to denote worlds. Please choose one (I find the latter clearer as it appears to correspond to the traces lambda_1, lambda_2 whose completion brings about those worlds).

In Sections 1 and 2, I'm not sure the authors mean "BDI". BDI is more abstract than a "programming paradigm" and the original literature does not, unless I am misremembering, say anything about pre-defined plans (procedural knowledge). From the examples listed, I think that they are referring to PRS systems (which are based on BDI, but not everything BDI is PRS).

It is a trivial point, but I'm not convinced about the use of the word "dynamics". The paper refers to "changing" dynamics and "constant" or "fixed" dynamics, but the latter sounds like a contradiction in terms. Even if it could be read as "changing in a constant way", this does not appear to be what the authors mean in practice. They refer to their specific contribution as a "dynamic confidence measure" (Section 6), but given their use of the term, I'm not sure if they mean a "constant dynamic confidence measure" or a "dynamic dynamic confidence measure"!

The text is well written, with only a few typos:
- Section 1, paragraph 1: "leveraging on" -> "leveraging"
- Section 2, paragraph 4: "this selection then is described" -> "this selection is described" (or "is then")
- Section 2, paragraph 5: floating comma before "lessening the burden"
- Section 3, paragraph 14: "relative its last n execution" -> "relative to its last n execution"
- Section 4, paragraph 2: "comprise of" -> "comprise"
- Section 5, Experiment 2: "stetting" -> "setting"

Summary of review
While some concepts need to be explained more clearly, e.g. stability, I found this work interesting, convincing, well explained and well evaluated.

"Author Response (optional)
Authors response by Dhirendra Singh (2010-12-02 03:28:24) to review
Review Evaluation: Very helpful

Stability states how well-informed decisions were. Given all decision paths, success is rare and failure frequent. Upon success, we consider the decisions well-informed (stability/degree=1); but upon failures, we want to know how well-informed (degree=?). For threshold please read response to reviewer1.

Programmed context conditions are used as the first filter followed by learning filter to decide final applicability. Interpret "real" context conditions to mean the sum result of this filtering.

RecordDegreeStability "saves" d wrt P1,W1 so is not required to be passed on recursively.

Plans for G(r,5,s) configure all modules by: configuring module5 then posting and waiting for G(r,4,s') to finish. "Possible requests" and "overall response" both mean [-1.0:+1.0] in steps +/-c.

We agree that use of terms "BDI" and "dynamics" could be improved.