UAI2015
31th Conference on Uncertainty in Artificial Intelligence
July 12-16, 2015, Amsterdam, The Netherlands

Reviews For Paper

Paper ID 130
Title A Markov Game Model for Valuing Player Actions in Ice Hocke

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	The paper presents an analysis of a dataset of ice hockey events from real professional league games over a period of 7 years. The events in the dataset are used to define a Markov model that represents how the game progresses through time. Q-learning is used to estimate the values of states/actions in the Markov model. The idea is that this can give more detailed estimates of impact of actions on temporal events in the near and distant future. The model is examined to see how different contextual situations change the value of a state, and significant variance is found across most actions. The model also is used to look at temporal progressions of events. Individual players are also examined and some interesting things are noted.
Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize that novelty can sometimes be relative, and	A novel model for ice-hockey as a Markov game. In that sense it is novel, but as an application of existing techniques to a new dataset. A fair bit of engineering and development had to go into the process, though, so its more than just an off-the-shelf application.

we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	
Novelty numeric score	4-Worthy contributions, but not surprising
Technical quality. Are the results technically sound? Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did	A few things were unclear to me in this paper. - p7. last paragraph of "impact on scoring the next goal" - I am not clear on what the significance of these entropy numbers are. - p9. I found table 9 to be hard to interpret. It might make more sense to present the actual values rather than (or as well as) the differences. Its not clear what a difference of -0.019 means. Is this significant? Should be discussed at least. - how were the player scores computed? The paper seems to focus entirely on teams - there is no mention of players in the state-space sections at all, then suddenly we get tables 6 and 7 showing individuals. I guess figure 2 is averaged over all players? This needs to be clarified.

the authors clearly assess both the strengths and weaknesses of their approach?	- Why are the player positions not considered at all? There are a lot more centermen in table 6 than table 7 for example, so that could be useful feature.
Technical quality numeric score	4-A paper that may be strong in other respects, but not technically.
Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful 5 or 10 years from now? Is this work that can be built on by other researchers?	The paper is well written and reasonably clear (although some elements are not - see below). It does a fairly good job of showing some interesting new results. It was somewhat difficult to see the significance of the results in the end, however, as there was nothing really to compare against. Some elements of player performance (e.g. comparing to points) were fairly obvious, and so I came away from the paper wondering how useful this would be in a practical situation. If the same model were able to make predictions about games or teams, then the paper would be much stronger.
Significance numeric score	4-Reasonable contribution to a minor problem
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? It is the responsibility of the authors of a paper to write clearly.	3-Quality of writing is good, but could be improved with some editing

rather than it being the duty of the reviewers to try to extract information from a poorly written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	5-A good paper overall, accept if possible. I vote for acceptance, although would not be upset if it were rejected because of the low acceptance rate.
(Optional) Additional Comments to the Authors: please add any additional feedback you wish to provide to the authors here. For example, if the quality of writing in the paper is not excellent, please provide some feedback to the authors on how the writing can be improved.	minor things: p1, top of second column: should "policy-on" be "on policy"? p4 "findings" 3 and 4: one says "the probability" while the other is the "conditional probability". they should both be conditional and the paper should state on what (being on the powerplay I assume) p4 "findings" 5: I'm not clear why the 64.8% goal differential translates into more short-handed goals - short-handed is sometimes written shorthanded - be consistent - define short-handed goal and powerplay goal in 3.1

	p5 1, 325, 809 should not have spaces after each comma p5 I did not find figure 1 to be terribly helpful, as there are many things going on which are not explained. For example, what is R(s) and why does it increase to 1 when there is STOPPAGE, but not on the goal? This figure should be simplified and/or explained in more detail. Perhaps use as a running example in the text to make things more clear? p8. In figure 2 what are the red dots and green stars? I think the green "asterisks" are defined (Without mentioning the color), but red dots are presumably the data lying outside 2.7sd? What are teh dashed lines? p7. in "impact on scoring the next goal", there are a couple of examples given, but its really hard to figure out where those are in figure 2, so more explanation could be useful here 11 sections seems like a lot for a 10-page paper. PErhaps combine some and make subsections? e.g. sections 5-10 could be merged into section 4

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	The authors model hockey as a Markov game and learn a model of it to study the impact on key events (goals and penalties) of local actions by individual players. Their results show correspondence to intuitive and empirical analyses of the game. They also offer the potential for deeper analyses that might provide additional insights into the game.

Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize that novelty can sometimes be relative, and we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	This is a novel application of Markov games as a model of professional hockey. While there are certainly many statistical analyses of the sport, this a highly UAI-relevant approach that appears very promising.
Novelty numeric score	6-Several novel and surprising contributions
Technical quality. Are the results technically sound? Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out	The authors' approach appears sound, although their individual modeling decisions are not always well-motivated. For example, why is the reward tied to directly to penalties, and not tied to only goals? Penalties have no inherent cost, although they have an expected negative reward in that the shorthanded team is more likely to give up a goal. I did not see any motivation why an additional reward component for penalties was included.

and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did the authors clearly assess both the strengths and weaknesses of their approach?	Section 11.2's evaluation of the player valuation would be much more compelling if the authors aggregated goal impact scores across each team and compared it against the team's records. The objective of this work is presumably to provide a metric for hockey performance that is more consistently informative than current sabermetric and scouting techniques. Therefore, showing that the players with the highest impact scores are all highly-regarded players is a good sanity check, but it does not demonstrate additional value over existing techniques. What would be interesting would be to see how team performance was affected by individual impact scores. Maybe the authors could look at team performance when individual players were missing. They should also look to see whether impact scores are consistent from year to year, and are thus indicative of player quality in a way that might be resistant to changes in context (e.g., different teammates).
Technical quality numeric score	5-Technically adequate for its area, solid results
Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful	The authors face a serious challenge in trying to demonstrate that their model provides a surprising insight. On the one hand, if their model matches conventional wisdom, that is a validation of their model, but not a demonstration of additional value. On the other hand, if their model challenges conventional wisdom, the onus is on them to demonstrate that conventional wisdom is wrong. The latter case would be a very compelling demonstration of significance, but it is obviously very hard to do. It is certainly suggestive that, for example, Jason Spezza has a good impact score, but bad +/-, but

5 or 10 years from now? Is this work that can be built on by other researchers?	I think it is already accepted that +/- is highly dependent on team context. Basically, the authors need to dig around to find a stronger example of this kind of surprising finding, in which case this paper becomes highly significant. The existing Markov process models have a clear limitation by not being able to represent actions. However, it would still be interesting to know whether their valuation of states is similar to your own. Similarly, how does your valuation compare to the regression models? It would be very valuable to see how the richer MDP representation contributes to differences in the end model. Section 5.1 (side note: there is no 5.2) is helpful in this regard, in that it provides a qualititative comparison of outcomes in your model against earlier findings. But perhaps there is some quantitative comparison that could be done as well?
Significance numeric score	5-Solid contribution to relevant problem
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? It is the responsibility of the authors of a paper to write clearly, rather than it being the duty of the reviewers to try to	3-Quality of writing is good, but could be improved with some editing

extract information from a poorly written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	6-A very good paper, should be accepted. I vote and argue for acceptance, clearly belongs in the conference
(Optional) Additional Comments to the Authors: please add any additional feedback you wish to provide to the authors here. For example, if the quality of writing in the paper is not excellent, please provide some feedback to the authors on how the writing can be improved.	Section 4: "This reflect" should be "This reflects"

Masked Reviewer ID: Assigned_Reviewer_5

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper	This paper develops a model of NHL hockey that tries to solve the credit assignment problem using reinforcement learning techniques so that the value of

is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	players can more accurately be measured. A large scale dataset of play-by-play data is used, a simple model is developed, and results are reported. Overall it's an interesting idea, but I don't think this paper is a perfect fit for UAI. For the UAI audience, I don't think the particular sport is important. So related work should not just focus on Hockey, but it should also include basketball and other similar sports. In this case, there is related work that uses similar ideas by Cervone et al [A]. This should be cited and discussed. The Cervone et al approach focuses on building a real-time "expected possession value" model, then attributes changes in EPV to specific players. It also uses fairly sophisticated spatial models of the coure and more detailed information about the games. The modelling in this submitted work is rather strange, as state is stored as a concatenation of all previous actions, and (assuming I'm interpreting the notation correctly) the dynamics model assumes that there is a unique transition probability for all (context x history, context' x history') pairs. So am I correct in assuming that Occ(s,s') is almost a functional relationship, i.e., there is only one observed s' for the majority of s? I'm then left having difficulty with the intuition of how credit can reliably be assigned. This seems to me to be an under-determined problem. Overall, I think this is an interesting application, but the methodology used is

	quite simple and I'm not sure would be particularly interesting to the UAI audience. The core idea of using something like Q values to aid in the credit assignment problem in evaluating sports player impact is interesting, but as discussed above has appeared in previous work. A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes Daniel Cervone, Alex D'Amour, Luke Bornn, Kirk Goldsberry http://arxiv.org/pdf/1408.0777v1.pdf (previous version at Sloan Sports analytics conference: http://www.sloansportsconference.com/wp-content/uploads/2014/02/2014_SSAC_Pointwise-Predicting-Points-and-Valuing-Decisions-in-Real-Time.pdf)
Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize that novelty can sometimes be relative, and we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is	See above. Key ideas have appeared in previous work.

novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	
Novelty numeric score	3-Minor variations to existing techniques
Technical quality. Are the results technically sound? Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did the authors clearly assess both the strengths and weaknesses of their approach?	The presentation appears technically correct, but modelling state as a concatenations of actions seems overly simplistic to me in this setting.
Technical quality numeric score	3-Claims not completely supported, assumptions or simplifications unrealistic

Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful 5 or 10 years from now? Is this work that can be built on by other researchers?	The problem is of interest to a very wide audience (sports fans), but I'm skeptical that the approach in this paper will impact mainstream sports.
Significance numeric score	4-Reasonable contribution to a minor problem
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? It is the responsibility of the authors of a paper to write clearly, rather than it being the duty of the reviewers to try to extract information from a poorly	3-Quality of writing is good, but could be improved with some editing

written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	3-A weak paper, just not good enough. I vote for rejecting it, but could be persuaded otherwise.