| Subject: | FW: ECML-PKDD 2012: invitation to submit extended version of your paper to MLj special issuen |
|---|---|

------------------------------------

## Reviews For Paper

| | |
|---|---|
| **Track** | 1. Research track |
| **Paper ID** | 297 |
| **Title** | Score-based Bayesian Skill Learning |

**Masked Reviewer ID:** Assigned_Reviewer_1

**Review:**

| Question | |
|---|---|
| Give your own brief summary of the paper: | This paper describes an extension of the score model TrueSkill. The extension includes modeling offence and defence skill separately, as well as the interaction of these skills for score-based match outcomes generation. The presented results are relevant in ranking, gameplay managing, sport betting and bookmarking. |
| State three strong points of the paper: | 1) Technically well justified and presented. 2) Empirical improvements with small training sets. 3) Empirical evidence with real data. |
| State three weak points of the paper: | 1) Quite limited/focused application scenario. 2) Does not consider time-dependent skill modeling (probably too much for a conference paper anyway). 3) No results expressing the hardness/complexity of time/space of the proposed methods in comparison with existing ones. |
| Indicate to what extent the paper has the potential to inspire the research community by introducing new and relevant problems, concepts, solution strategies, and ideas: | Above average |
| Indicate to what extent the paper contributes to solving a problem widely | Somewhat |

| Question | |
|---|---|
| recognized as both challenging and important: | |
| Indicate to what extent the paper addresses a novel area of impact of machine learning and data mining: | Above average |
| Give detailed comments justifying your evaluation of the paper, covering scientific rigour and correctness, challenges overcome, quality and reproducibility of the experiments (if any), and presentation: | This paper is well written and presented. The technical results, namely Lemma 1 is quite interesting. The challenge of improving score-based models is relevant and the proposed methods outperformed existing ones for relevant and representative datasets. For this reason I think the paper is suitable for publication in this conference.<br><br>Minor typos:<br><br>- pag 7, Eq 9 there is a missing full stop. |

**Masked Reviewer ID:** Assigned_Reviewer_3
**Review:**

| Question | |
|---|---|
| Give your own brief summary of the paper: | An existing Bayesian model for skill learning in games, TrueSkill, is extended in two directions: (a) scores are modeled in addition to plain win/lose/draw outcomes; (b) skills are modeled in terms of separate defence and offence skills. Experiments on selected real data sets demonstrate the potential advantages of the proposed model. |
| State three strong points of the paper: | 1. Sound work (clear problem, model, evaluation).<br>2. The presentation is sufficiently polished.<br>3. Relevant literature adequately cited. |
| State three weak points of the paper: | 1. Rather incremental, routine applied modeling work with no real methodological innovations.<br>2. The importance of the application (skill learning from game data) for machine learning research is somewhat questionable.<br>3. The observation that the proposed extended model performs better than TrueSkill on small data but not as much on larger data is counter intuitive and deserves a more thorough discussion. |
| Indicate to what extent the paper has the potential to inspire the research | Somewhat |

| | |
|---|---|
| community by introducing new and relevant problems, concepts, solution strategies, and ideas: | |
| Indicate to what extent the paper contributes to solving a problem widely recognized as both challenging and important: | Somewhat |
| Indicate to what extent the paper addresses a novel area of impact of machine learning and data mining: | Somewhat |
| Give detailed comments justifying your evaluation of the paper, covering scientific rigour and correctness, challenges overcome, quality and reproducibility of the experiments (if any), and presentation: | Section 4.2 uses over 2 pages to develop a variational routine for inference in the non-gaussian case. I am wondering whether the variational method is justified in this case. Namely, it seems like the functions are very low-dimensional (one or two) and therefore numerical integration methods and piecewise representations should be more accurate and sufficiently fast to compute with.<br><br>While the presentation is fairly good, there is still room for improvement; examples:<br><br>- If "p" and "d" denote variables/parameters, then it is better to refer to the probability model (measure) by, say, "P" and to the integration measure by non-italic "d".<br><br>- In Eq.(3) should write "\|d\| < epsilon" instead of "d < \|epsilon\|". (?!)<br><br>- Page 7: "Figure 2(N)". What is N referring to here?<br><br>- It seems like the first occurence of variables x and y is on page 7. What are they referring to? What are they dimensions?<br><br>- Page 13: "models except AFL" should be "data except AFL". (?!)<br><br>- Figure 4: Label the curves by symbols (large enough circles, crosses, triangles, ...) that are easy to distinguish from each other. |

**Masked Reviewer ID:** Assigned_Reviewer_4
**Review:**

| Question | |
|---|---|
| Give your own brief summary of the paper: | Extensions to a Bayesian skill learning algorithm (TrueSkill) is presented that also take into account the actual scores of a game (rather than just win-loss-draw) in modelling the skill levels of the individual players/teams. Three variants of the original TrueSkill model are discussed, two of which are based on separate models of offense and defense skills of the players. Experimental results on 3 real-world datasets demonstrate some improvement over TrueSkill. |
| State three strong points of the paper: | . interesting ideas for extending the original model<br>. informative experimental results |
| State three weak points of the paper: | . rather specialised problem => limited impact/significance for general machine learning<br>. paper is hard to follow for non-experts in probabilistic modelling (in particular, the presentation of the inference algorithms is rather dense) |
| Indicate to what extent the paper has the potential to inspire the research community by introducing new and relevant problems, concepts, solution strategies, and ideas: | Somewhat |
| Indicate to what extent the paper contributes to solving a problem widely recognized as both challenging and important: | Somewhat |
| Indicate to what extent the paper addresses a novel area of impact of machine learning and data mining: | Somewhat |
| Give detailed comments justifying your evaluation of the paper, covering scientific rigour and correctness, | The paper addresses a rather specialised problem, but does so in what seem to me to be creative and reasonable ways. Thus, the impact on general machine learning research is probably limited. However, the problem itself is definitely interesting. I am not a top expert in probabilistic modelling (and, in this respect, probably representative of some part of the ECML |

| | |
|---|---|
| challenges overcome, quality and reproducibility of the experiments (if any), and presentation: | community), and I found it hard to follow some parts of the paper. While the modelling part is generally readable, the inference part is very dense. Since the description of the Poisson-OD update operator is disproportionately longer and more detailed than the others and relies on "Supplemental material" anyway, I would suggest shortening this part (drop the detailed proof of Lemma 1 and present the fixed-point algorithm more directly; move some of the dropped stuff to the "supplemental material" part) and using the saved space for some more detailed explanation of the terms in the section 2 (see below).<br><br>Comments on Section 2:<br>It would help the reader if some of the concepts were explained a bit more explicitly. In particular, explain what "performance" is (a varying skill level, presumably); what the $\delta$ function is (in the definition of the "performance differential"); what the II function in equation (3) is.<br>Also, the sentence that introduces the skill level vector $\mathbf{l}$ (first paragraph on p.3) is unclear: you say that l is in R^n, but at the same time say "the skill level vector l .. of the teams in M (namely l_i and l_j)", which implies that l would only be a two-component vector. Please rephrase.<br><br>Comment on Section 3:<br>The Gaussian models of "performance" (second row of the figure) mention means o_i and d_j (presumably meaning "offense skill" and "defense skill", respectively). However, these concepts have not yet been introduced at this point -- they only enter the scene later, in section 3.1. I think Fig. 1 should still refer to skill levels l_i, l_j instead. |

## Meta-Reviews For Paper

| | |
|---|---|
| **Track** | 1. Research track |
| **Paper ID** | 297 |
| **Title** | Score-based Bayesian Skill Learning |

**Masked Meta-Reviewer ID:** Meta_Reviewer_1

**Meta-Reviews:**

| Question | |
|---|---|
| Brief motivation of your recommendation to the PC chairs. Please refer to the 3 overriding criteria of acceptance from the CfP in particular (reproduced above). Papers that | Solid contribution to a problem with immediate practical utility.<br>Well written although dense in certain points.<br>Slight negative aspects are somewhat limited problem setting and incrementality compared to earlier work.<br>Overall a very nice paper. |

| excel on one of these criteria should be given preference over papers that are average in all. | |
| --- | --- |