The MultiSkill Tennis Model

Estimating player skills on serve and return with dynamic Bayesian networks

Daniel Korzekwa (daniel.korzekwa@gmail.com); Machine Learning Summer School, Tübingen Germany, 2013

1. Abstract

Popular models for estimating skills in tennis, chess and Xbox online games, such as Elo, Glicko and TrueSkill, characterise each player using a single skill variable, rather than differentiating between skills on attack and defence. These models also assume that all games take place under the same conditions, such as weather or tournament rank. These limitations can lead to poor predictions of expected game outcomes. This poster presents an overview of a new model for estimating skills in tennis based on dynamic Bayesian networks that takes into account both multiple skill sets and different game conditions.

2. The MultiSkill tennis model

The proposed MultiSkill model takes inspiration from the TruSkill[1] and TrueSkill Over Time[2] rating models, and applies three main approaches: (1) using Gaussian variables to model the evolution of player skills over time, (2) representing games with Bayesian networks and (3) using expectation propagation[3] inference to update skills given the observed outcome of a game.

In contrast to single-skill-based TrueSkill, MultiSkill models each player with multiple skills that vary depending on the conditions of a particular game. For example, tennis players exhibit different skills in terms of their abilities to serve, return and score aces. These multiple skills of a single player can intuitively be considered as the combination of the single skills of multiple individual players forming a team in TrueSkill, except that they are not necessarily additive.

Moreover, unlike in TrueSkill, player skills are updated after every point of a game in MultiSkill, with important points, such as break points, contributing more to the updates. Nowadays, historical point-by-point data is available in tennis that allows for modelling player skills on a finer level of granularity.

For illustration, consider an example of updating the skills of players P_1 and P_2 serving and returning a ball, respectively, after a single tennis point is played (Figure 1). The mean skill is expected to increase for the point winner and decrease for the loser. Note that the skill variance is usually moderated by observations of new information.

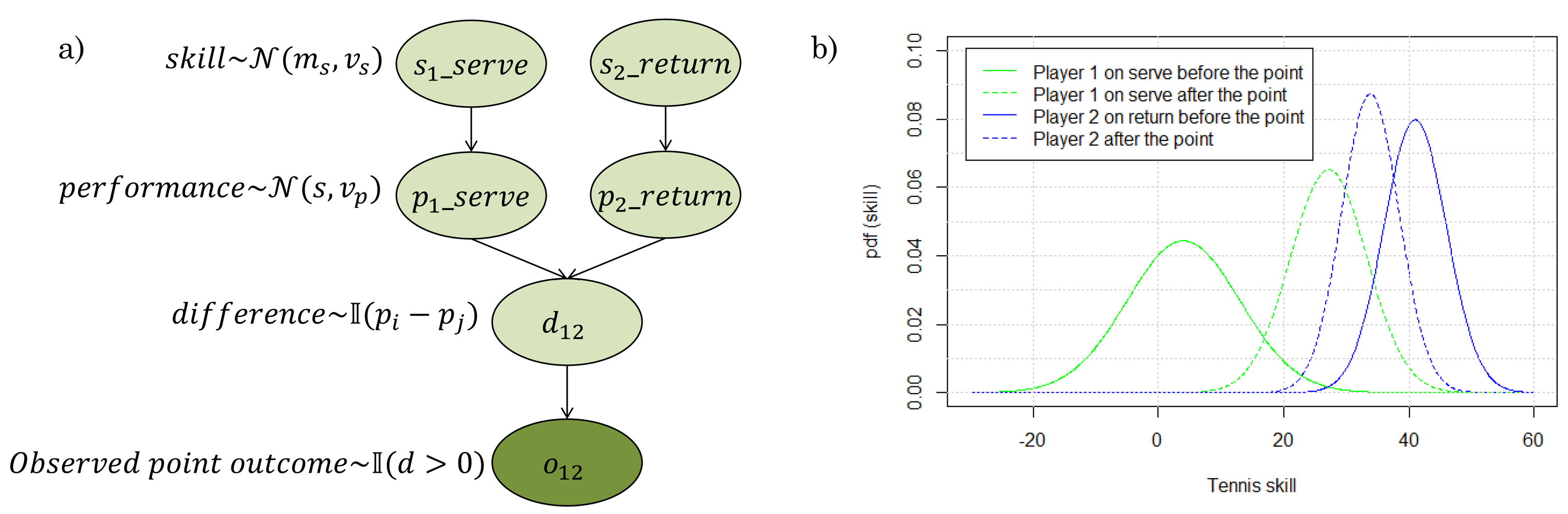

Figure 1: (a) Bayesian network for a single tennis point. (b) Skills of players P_1 and P_2 serving and returning a ball, respectively, before and after player P_1 wins a point.

Historical matches over time form a dynamic Bayesian network[4] (Figure 2). Such a network firstly allows for learning network parameters, such as performance variance, with the expectation maximization algorithm[5,6]. Secondly, estimating skills with a smoothing technique is more accurate than using a simple online Bayesian update.

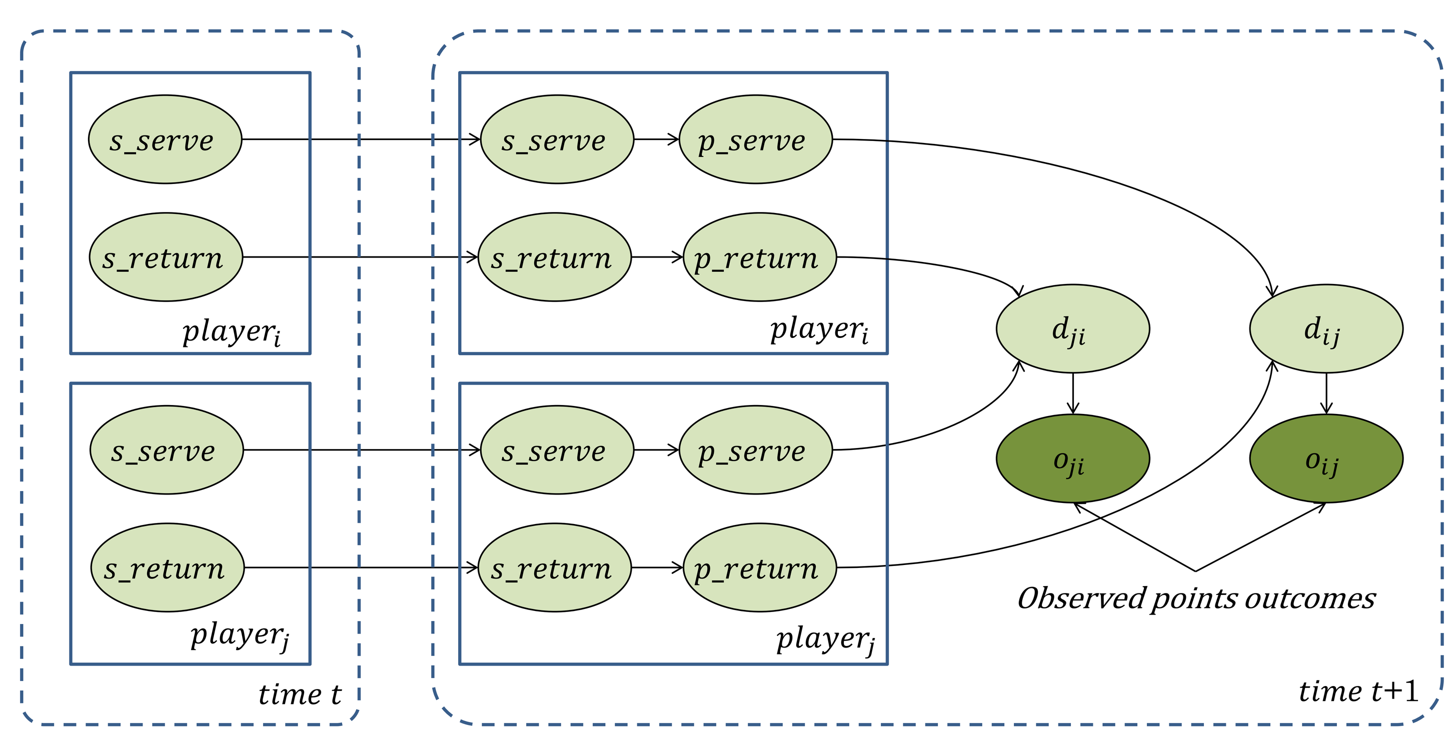

Figure 2: Dynamic Bayesian network for multiple players playing many tennis points over time.
s – skill, p – performance, d – skills difference, o – outcome of tennis point (win/lose).

3. Predicting the outcome of a tennis match

A hierarchical Markov chain[7] is set up for a tennis game, tie-break, set and match (Figure. 3). Its inputs include the probabilities of both tennis players winning a point on serve and return, and the context of a game, such as the current score or tournament rank. Point probabilities are derived from skills on serve and return with an expectation-propagation inference algorithm.

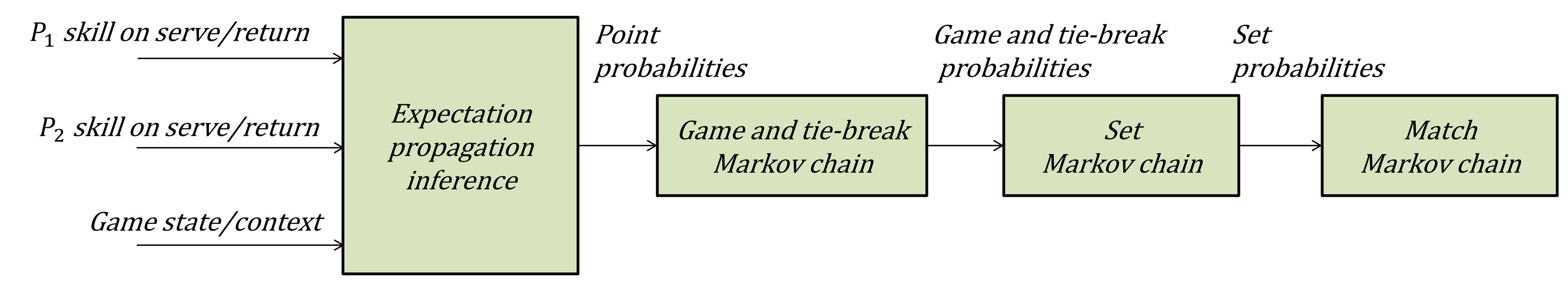

Figure 3: Hierarchical Markov chain for a tennis match.

4. MultiSkill vs TrueSkill (online learning scheme)

The MultiSkill and TrueSkill models generally predict the outcomes of tennis matches quite well (Figure 4a), with the MultiSkill model being slightly more accurate (Table 1) and also converging faster for unknown players (Figure 4b).

	2007	2008	2009	2010	2011	2007-2011
MultiSkill	33.50%	35.70%	32.65%	35.55%	31.06%	31.73%
TrueSkill	36.09%	40.17%	37.26%	38.02%	36.28%	32.93%

Table 1: Fraction of players predicted incorrectly as winners (annual averages of 1445 and 196 players)

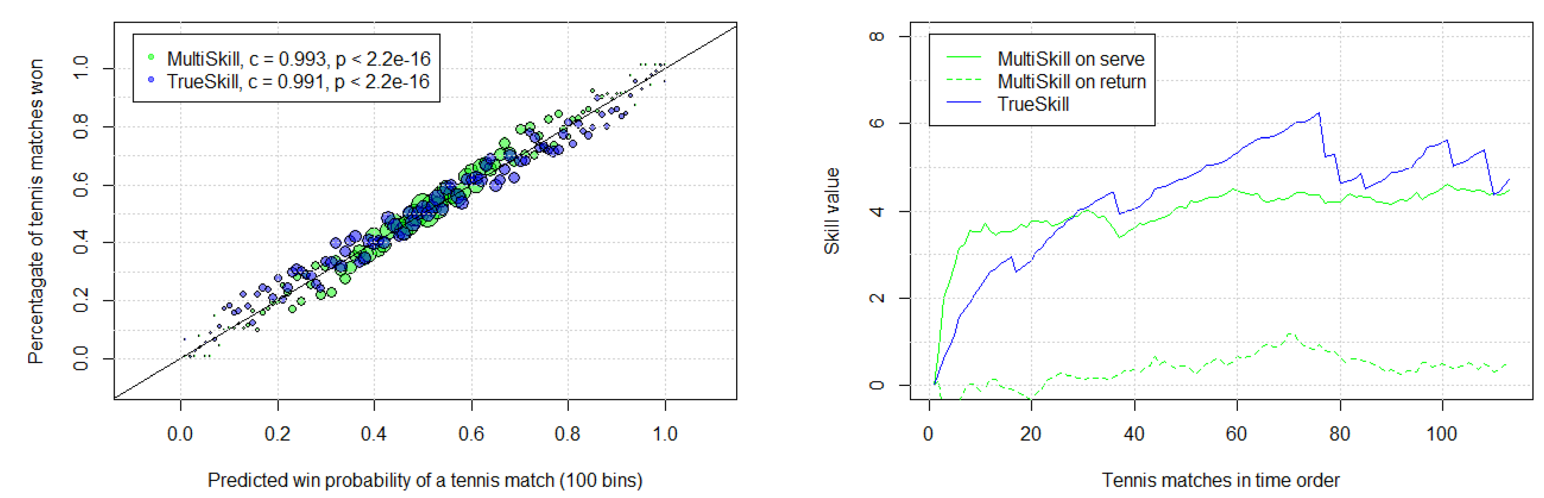

Figure 4: (a) Predicted vs expected match probabilities for 7226 ATP tennis matches during 2007–2011. (b) Convergence speed of player skills for Roger Federer in 112 tennis matches during 2006 and 2007.

Both rating models are consistent with the implied probabilities provided by the Betfair betting exchange[8] as illustrated in Figure 5b; however, MultiSkill adapts rapidly to changes in player skills over time. For example, TrueSkill overestimates the probability of Roger Federer winning the semi-final match against Novak Djokovic during the 'Masters 1000 Indian Wells 2011' tournament (Table 2, Figure 5b), which is simply due to TrueSkill not considering Federer's poor performance at the point level during the first five rounds of the tournament.

	Federer win probability
MultiSkill	0.42
TrueSkill	0.52 (overestimated)
Betfair Exchange	0.43

Table 2: Federer vs Djokovic in semi-final of Indian Wells 2011 tournament

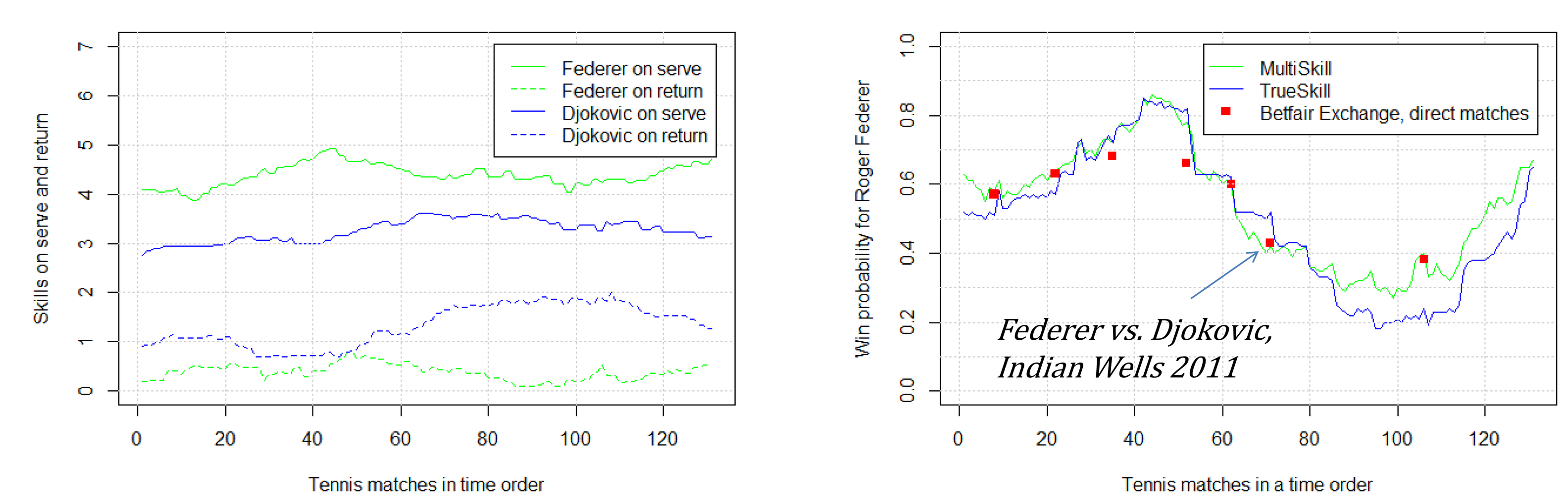

Figure 5: Analysis results for Roger Federer vs Novak Djokovic in 131 tennis matches during September 2010 to December 2011. (a) MultiSkill skill levels. (b) Probabilities of winning a match according to MultiSkill, TrueSkill and the Betfair betting exchange.

5. Conclusions

Estimating multiple skills of a tennis player provides deeper insights into his or her performance compared to modelling each player with a single skill only. This approach increases the accuracy of predictions of outcomes for tennis points, games, sets and matches.

Nonetheless, there are still several tennis matches for which there are considerable discrepancies between MultiSkill and Betfair betting exchange predictions. Such discrepancies could be due to fraud or inefficiency in the Betfair betting exchange markets, or to the MultiSkill model not considering certain factors influencing the winning probabilities in tennis that are known to the people trading in the Betfair betting exchange.

6. Future work

Future work will include the addition of further skills for tennis players to the MultiSkill model, such as scoring aces, physical endurance and stress resistance, which might break the following transitivity rule in the pairwise comparison model:

Player 1 is better than Player 2
Player 2 is better than Player 3
Player 1 ***is not better*** *than Player 3*

We also intend to analyse the importance of contextual factors in predicting the winner of a tennis match, including the weather condition, tournament rank, time of the year, player's age and prize money.

7. Model implementation

Model component	Software
Bayesian learning and inference	bayes-scala
Tennis-model implementation and analysis for MultiSkill and TrueSkill models	tennis-player-compare
Tennis hierarchical markov chain	tennis-probability-calculator

Table 3: Software for the MultiSkill tennis model (available on github)

8. References

1. Ralf Herbrich, Tom Minka, Thore Graepel. TrueSkill TM: A Bayesian Skill Rating System, MIT Press, January 2007
2. Pierre Dangauthier, Ralf Herbrich, Tom Minka, Thore Graepel. TrueSkill Through Time: Revisiting the History of Chess, MIT, 2008
3. Thomas P Minka. A family of algorithms for approximate Bayesian inference. PhD Thesis, MIT, 2001
4. Daphne Koller, Nir Friedman. Probabilistic Graphical Models, Principles and Techniques, MIT Press, 2009
5. Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2009
6. Zoubin Ghahramani, Geoffrey E. Hinton. Parameter Estimation for Linear Dynamical Systems, Technical Report, 1996
7. Tristan J. Barnett. Mathematical Modelling In Hierarchical Games with specific reference to tennis, PhD Thesis, 2006
8. Betfair betting exchange, http://www.betfair.com/exchange

9. Acknowledgements

I thank to Thore Graepel, Ralf Herbrich, Franc Klaassen and Tom Minka for their valuable comments on my work.