Design and Secure Evaluation of Side-Choosing Games

Ahmed Abdelmeged, Ruiyang Xu and **Karl Lieberherr**
CCIS/PRL
Northeastern University, Boston, MA 02115
email: lieber@ccs.neu.edu

Abstract

We present an important, general class of new games, called side-choosing games (SCGs), for "gamifying" problem solving in formal sciences. Applications of SCGs include (1) peer-grading in teaching to (2) studying the evolution of knowledge in formal sciences to (3) organizing algorithm competitions. We view SCGs as a new and general model for formulating formal problems that need to be solved using human computation. Our interest in this paper is on how to evaluate an SCG tournament fairly and effectively. We observe that a specific kind of collusion, where players lie about their strength and sacrifice themselves, could bias the evaluation of SCG tournaments dramatically. Following the idea of Social Choice Theory in the sense of Arrow, we take an axiomatic approach to guarantee that a specific kind of collusion is impossible. We prove the Collusion-Resistance Theorem as a general principle for designing collusion-resistant evaluations for SCG tournaments. The Collusion-Resistance Theorem is surprising: it tells us to be *indifferent* to wins but to count certain kinds of losses for scoring players and ranking them. If collusion is not an issue, we offer a family of useful ranking functions which are not collusion-resistant.

1 Introduction

A side-choosing game (SCG) $H = \langle G, GS, Q, p_x, p_y\rangle$ is based on an extensive form two-player game G between players p_x and p_y with perfect information and without ties, i.e., there is always a winner and a loser[1]. G is a game between two players, *white* and *black*. GS is a game state of G (i.e., a node of the game tree of G). Q is a proposition on the game state GS of the form: *white* or *black* has a winning strategy when *white* or *black* moves first.

The players p_x, p_y of an SCG have their preferred, static side (*white* or *black*), depending on whether they believe Q or !Q to be true. The players are free to choose their static side before the game. But during the game the players must have opposite "run-time" sides which we implement by making (per game) at most one of them the devil's advocate (or forced).

The side-choosing game $\langle G, GS, Q, p_x, p_y\rangle$ produces a game result row consisting of (1) the **winner** (p_x or p_y) (2) the **loser** (p_x or p_y) and (3) at most one **forced** player (p_x or p_y or 0, if none was forced). A set of game results produced by multiple binary SCGs is called an SCG-Table.

1.1 Examples

Consider the chess position GS in Fig. 1 as a side-choosing game[2]. The game G is chess, modified so that winning for white means to mate the black king in 2 moves. The proposition Q says that white starts in GS and wins. We have two players, Alice and Bob, who study G and GS and make their side choices. Alice believes she can win as white (she "sees" the mate in two) and therefore her side choice is *white*. Bob does not see the mate in two and therefore he wants to be *black*. The game is played and Alice wins (how is left as an exercise to the reader; there is only one optimal move for white.). Game result row= (**winner**="Alice", **loser**="Bob", **forced**=0). If *white* plays b3 and *black* f3 then *white* mates with Qb2. *White* wins but only because *black* made a mistake. The correct move for *black* is c4 (not f3) and *white* cannot mate in the next move. This example gives the wrong impression that playing one perfect game reveals the winning strategy (solution). But this is not the case most of the time.

Next we consider a large family of examples of side-choosing games: the family consists of semantic games (Kulas and Hintikka 1983) for claims with side choice added. The game G is defined by an interpreted logical sentence between white (proponent, existential quantifier) and black (opponent, universal quantifier). The outermost quantifier of the sentence determines who moves first. For white, the game is about making the sentence true by assigning values to variables. Black tries to prevent this. Side-choosing games exist for many different logics such as first-order predicate logic, higher-order logics and independence-friendly logic. A first important subfamily are claims in formal sciences with side choice added. A second important subfamily are claims related to algorithm specifications for algorithm com-

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1]Extensive form games are widely used to model multi-agent sequential decision making. They are represented by game trees.

[2]Taken from the collection: SEVENTY-FIVE CHESS PROBLEMS by John Thursby, Trinity Coll., Cambridge, 1883.

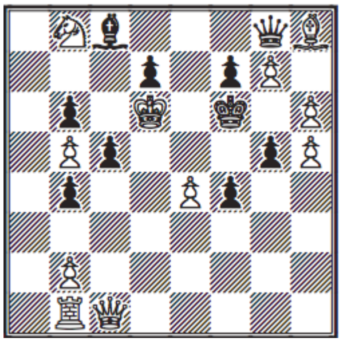

Figure 1: An SCG example: white mates in two

petitions (a la TopCoder, see topcoder.com or Kaggle, see kaggle.com).

1.2 Motivation

Claims are ubiquitous in human reasoning. An SCG can be viewed as a plausibility check of a claim. A game state GS of an extensive form game G is a model of a claim. When I argue that a claim is true but then I cannot defend it using a plausibility check, there must be something wrong with my argument. We are interested in how to aggregate multiple such plausibility checks in a robust ranking manner, considering incentive and trust.

There are two kinds of incentives in SCG: the incentive (1) to be top-ranked which brings money or fame and (2) to get feedback during game play which builds skills and provides opportunity for learning. Incentive (2) suggests productive applications of SCGs in education.

Trust in the SCG approach is related to the belief that good work as a player will get rewarded and that it is not possible to be top-ranked without doing good work. There should be no sneaky ways to game the system: money or fame must be well deserved. Trust can be broken in at least two ways: (1) by defining games but not checking that all game rules are perfectly followed and (2) by having tournaments and evaluations where you can succeed without hard work. Point (1) is addressed by having reliable software to check the game rules related to the claims. This paper is addressing point (2). One important consideration in our ranking approach is to prevent sybil attacks. In an online competition, several sybils might enter and help others to win thereby preventing the strong players to win. In the presence of collusion-resistance, sybils have no effect on overpowering perfect players.

This workshop paper is a very condensed version of the main ideas in the dissertation (Abdelmeged 2014). (Abdelmeged, Xu, and Lieberherr 2015) gives a longer version of the material described in this paper, including complete proofs and more related work.

2 Main Theory

We discuss the ranking of players based on an SCG-table T under the axioms "winning cannot lower your rank" and "losing cannot increase your rank" and "games you don't control don't affect your rank". You are in control if you participate and are not forced.

The axioms are formalized by expressing that adding a row where player p_x satisfies a property, will keep the ranking of p_x with respect to other players p_y invariant.

Let P be the set of all players. $R(P)$ is the set of possible game results for P (for formal definition see (Abdelmeged, Xu, and Lieberherr 2015)). For our theory we define a few basic predicates: $\forall p_x \in P, \forall r \in R(P)$

$$participant(p_x, r) \Leftrightarrow p_x \text{ is a participant in the game } r$$
$$win(p_x, r) \Leftrightarrow p_x \text{ won the game } r$$
$$loss(p_x, r) \Leftrightarrow p_x \text{ lost the game } r$$
$$forced(p_x, r) \Leftrightarrow p_x \text{ is forced to choose a side in the game } r$$
$$control(p_x, r) \Leftrightarrow participant(p_x, r) \land \neg forced(p_x, r)$$
$$fault(p_x, r) \Leftrightarrow loss(p_x, r) \land \neg forced(p_x, r)$$

We also define counting functions for scoring players:

$$wf_{p_x}(T) = \text{the win count of } p_x \text{ in } T \text{ in a forced position}$$
$$wu_{p_x}(T) = \text{the win count of } p_x \text{ in } T \text{ in an unforced position}$$
$$lf_{p_x}(T) = \text{the loss count of } p_x \text{ in } T \text{ in a forced position}$$
$$lu_{p_x}(T) = \text{the loss count of } p_x \text{ in } T \text{ in an unforced position}$$

It's obvious that given table $T' = T \cup \{r\}$ and $X \in \{wf, wu, lf, lu\}$ the following transitional relations hold:

$$X_{p_x}(T') = \begin{cases} X_{p_x}(T) + 1 & \text{if X happens in } \{r\} \\ X_{p_x}(T) & \text{otherwise} \end{cases}$$

The equation above is critical to transform predicate logic axioms into algebraic formulas presented later in this paper.

2.1 Ranking Axioms

We define a pre-order $\preceq_U^T$ called the weakly better relation $\forall T \subseteq G$ based on the scoring function $U : \mathbb{N}\times\mathbb{N}\times\mathbb{N}\times\mathbb{N} \to \mathbb{R}$. Lower U means better player. For convenience, we drop the subscript and refer to it simply as $\preceq^T$.

We want to assign to each player a score solely based on the players' demonstration of ability. We use the above four counting functions, based on wins and losses and whether a player was forced, to calculate a player's score. We formally define the ranking relation as,

$$\forall p_x, p_y \in P, \forall T \subseteq R(P)[p_x \preceq^T p_y \Leftrightarrow U(wf_{p_x}(T), wu_{p_x}(T), wf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), wu_{p_y}(T), lu_{p_y}(T))] \quad (1)$$

We want the ranking relation to have the following properties defined in terms of table extensions:

- **NNEW**: Winning cannot lower your rank:

$$\forall p_x, p_y \in P, \forall T \subseteq R(P), \forall r \in \{r \mid r \in R(P) \setminus T \land win(p_x, r)\} [p_x \preceq^T p_y \Rightarrow p_x \preceq^{T\cup\{r\}} p_y] \quad (2)$$

- **NPEL**: Losing cannot increase your rank:

$$\forall p_x, p_y \in P, \forall T \subseteq R(P),\\ \forall r \in \{r \mid r \in R(P) \setminus T \wedge loss(p_y, r)\}\\ [\, p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (3)$$

- **CR**: Games you don't control don't lower your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P),\\ \forall r \in \{r \mid r \in R(P) \setminus T \wedge \neg control(p_x, r)\}\\ [\, p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (4)$$

2.2 Universal Domain

From equation 1, it is clear that for every logically possible game result table T, we have a valid preorder. This implies that our ranking relation satisfies the Universal Domain property.

2.3 Anonymity

From equation 1 it is clear that the scoring function ignores the identity of the player in calculating the score. Hence, the ranking relation $\preceq^T$ is unaffected by changing labels and therefore anonymous.

2.4 Monotonicity of U and Notations

As we score a player solely based on the player's wins and losses, NNEW and NPEL imply that the function U is monotonic. One interesting property of the parameters of U for a particular player is that when we add a new game to the existing game result table T, at most one parameter increments. This allows us to define the following notations:

$U \uparrow_x$: U is monotonically non-decreasing on the parameter x

$U \downarrow_x$: U is monotonically non-increasing on the parameter x

$U \wr_x$: U is indifferent on the parameter x

3 Properties of Ranking Relations

In this section, we reformulate the axioms as equivalent monotonicity constraints. It is easier to reason in the space of monotonicity constraints than in the space of predicate logic.

3.1 Collusion Resistance (CR)

Given $T' = T \cup \{r\}$, we reformulate CR as follows:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq\\ U(wf_{p_y}(T), wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T))\\ \Rightarrow\\ U(wf_{p_x}(T'), wu_{p_x}(T'), lf_{p_x}(T'), lu_{p_x}(T')) \leq\\ U(wf_{p_y}(T'), wu_{p_y}(T'), lf_{P_y}(T'), lu_{p_y}(T')) \quad (5)$$

Considering the definition of "not in control", there are 2 cases to treat:

I. Game results where p_x did not participate. Then p_y may have won or lost in a forced or unforced position against some third player p_z.

Let us consider the row $\{r\}$ where p_y wins over p_z in a forced position, given $T' = T \cup \{r\}$ we have,

$$U(wf_{p_x}(T'), wu_{p_x}(T'), lf_{p_x}(T'), lu_{p_x}(T')) =\\ U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T))$$

$$U(wf_{p_y}(T'), wu_{p_y}(T'), lf_{P_y}(T'), lu_{p_y}(T')) =\\ U(wf_{p_y}(T)+1, wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T))$$

From the CR constraint above, we have:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq\\ U(wf_{p_y}(T)+1, wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T)) \quad (6)$$

From equations 1 and 6, we get the monotonicity constraint,

$$U \uparrow_{wf} \quad (7)$$

Similarly, let us consider the case $\{r\}$ where p_y wins over p_z in an unforced position, given $T' = T \cup \{r\}$ we have,

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq\\ U(wf_{p_y}(T), wu_{p_y}(T)+1, lf_{p_y}(T), lu_{p_y}(T)) \quad (8)$$

From equations 1 and 8, we get the monotonicity constraint,

$$U \uparrow_{wu} \quad (9)$$

Using a similar argument, for the case where p_y loses over p_z in a forced position, we have

$$U \uparrow_{lf} \quad (10)$$

Also, for the case where p_y loses over p_z in an unforced position, we have

$$U \uparrow_{lu} \quad (11)$$

II. Game results where p_x is forced. In this case we have the following results:

$$U \uparrow_{wu} \wedge U \uparrow_{lu} \quad (12)$$

The full proof for (12) is in the technical report (Abdelmeged, Xu, and Lieberherr 2015). Now, CR can be summarized in terms of monotonicity constraints as,

$$U \wr_{wf} \wedge U \uparrow_{wu} \wedge U \wr_{lf} \wedge U \uparrow_{lu} \quad (13)$$

3.2 Non Negative Effect of Winning (NNEW)

Let us consider a game result $\{r\}$ where p_x won against a third player p_z. p_x could have won either in a forced or unforced position.

First, considering the case where p_x wins over p_z in a forced position, we have,

$$U(wf_{p_x}(T)+1, wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq\\ U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T), lu_{p_y}(T)) \quad (14)$$

From equations 1 and 14, we get the monotonicity constraint,

$$U \downarrow_{wf} \quad (15)$$

Similarly, for the case where p_x wins over p_z in an unforced position, we have

$$U \downarrow_{wu} \quad (16)$$

Summarizing the monotonicity constraints, we have,

$$U \downarrow_{wf} \wedge U \downarrow_{wu} \quad (17)$$

3.3 Non Positive Effect of Losing (NPEL)

Let us consider a game result $\{r\}$ where p_y lost against a third player p_z.

First, considering the case where p_y loses over p_z in a forced position, we have,

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq \\ U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T)+1, lu_{p_y}(T)) \quad (18)$$

From equations 1 and 18, we get the monotonicity constraint, $U \uparrow_{lf}$. Similarly, for the case where p_y loses over p_z in an unforced position, we have $U \uparrow_{lu}$. Summarizing the monotonicity constraints, we have,

$$U \uparrow_{lf} \wedge U \uparrow_{lu} \quad (19)$$

3.4 Local Fault Based (LFB)

As we want the ranking relation to satisfy all the three properties NNEW, NPEL and CR, from equations 13, 17 and 19, we get the monotonicity constraints,

$$U \wr_{wf} \wedge U \wr_{wu} \wedge U \wr_{lf} \wedge U \uparrow_{lu} \quad (20)$$

This tells us that the scoring function should be monotonically non-decreasing on faults and indifferent on other parameters. We call the ranking relation that uses a scoring function that satisfies equation 20 as Local Fault Based (LFB). The monotonicity constraints in equation 20 can be easily reformulated in predicate logic as follows. LFB: Games in which you don't make faults don't affect your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P), \\ \forall r \in \{r \,|\, r \in R(P) \setminus T \wedge \neg fault(p_x, r)\} \\ [\, p_x \preceq^T p_y \Leftrightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (21)$$

3.5 Collusion-Resistance Theorem

We just proved the Collusion-Resistance Theorem:

$$(\text{NNEW} \wedge \text{NPEL} \wedge \text{CR}) \Leftrightarrow \text{LFB}$$

This theorem tells us that collusion-resistant ranking functions have a simple form based on fault counting. There is an infinite family of such functions that can be used in the design of techno-social systems with guaranteed collusion resistance. The Collusion-Resistance Theorem is surprising: One would expect that counting wins against non-forced players would also be a good scoring function but it is not collusion resistant.

4 Conclusion

We propose the concept of side-choosing Game (SCG) as a model for plausibility checking of claims using a generalization of extensive form games. SCGs are useful for organizing techno-social systems for creating knowledge in Formal Sciences. Considering that a specific kind of collusion might compromise the truth, we modeled the ranking of participants functionally via three axioms or postulates: NNEW (Non-Negative Effect for Winning), NPEL (Non-Positive Effect for Losing) and the crucial axiom CR (Collusion-resistance, which says that games where one is not in control cannot affect ones ranking, hence preventing gaming the game). We prove the Collusion-Resistance Theorem which states that ranking has to be based on fault counting.

What comes next? Our plan is to deploy SCG-based applications on the web and gather the benefits of collective intelligence. So far, we have already applied SCG-based ideas and tools in designing courses at Northeastern University from algorithm and software development courses to basic courses on spreadsheets and databases. And we were planning to build a tool that can be used in MOOCs or algorithm competitions. An implementation of a domain-specific language for human computation in formal sciences is a challenge that requires several algorithms to be developed. Why not develop those algorithms with SCG-based human computation effectively bootstrapping the system based on user feedback. We view SCG as the programming language for human computation to solve complex problems.

Another important area that needs further work is where participants can propose new claims. A modular approach to solving claims is needed. For example, a complex claim C_1might be reducible to a simpler claim C_2 so that a solution for C_2 implies a solution for C_1. We propose a formal study of claim relations which can themselves be captured as claims and approached with side-choosing games.

Collusion is linked to trust in a tournament to find the best players. Collusion-resistance eliminates some collusion but there is still other collusion possible. We will report on this at the workshop.

Acknowledgments: This work was supported by the Northeastern University Lieberherr Fund supported by Pamela and Ian Holland and other alumni. We would like to thank Thomas Wahl for his feedback.

References: For space reasons we give only a few references. A complete set is in Ahmed Abdelmeged's dissertation (Abdelmeged 2014) on which this paper is based. However, side-choosing games are a contribution of this paper. Abdelmeged used semantic games with side-choice to formulate his results.

References

Abdelmeged, A.; Xu, R.; and Lieberherr, K. J. 2015. Design and Secure Evaluation of Side-Choosing Games. Technical Report NU-CCIS-TR-Nov-28-2015, Northeastern University. http://www.ccs.neu.edu/home/lieber/papers/side-choosing-games/wit-ec16-TR/secure-eval-WIT-EC-TR.pdf.

Abdelmeged, A. 2014. *Organizing Computational Problem Solving Communities via Collusion-Resistant Semantic Game Tournaments*. Ph.D. Dissertation, Northeastern University, Boston, MA, USA. Advisor-Karl Lieberherr.

Kulas, J., and Hintikka, J. 1983. *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*. Synthese Language Library. Springer.