Success chances in argument games: a probabilistic approach to legal disputes.¹

Régis RIVERET ^a, Antonino ROTOLO ^a, Giovanni SARTOR ^b, Henry PRAKKEN ^c, Bram ROTH

 ^a CIRSFID, University of Bologna, Italy
 ^b CIRSFID, University of Bologna and European University Institute, Law Department, Florence, Italy
 ^c Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands

Abstract. The outcome of a legal dispute, namely, the decision of its adjudicator, is uncertain, and both parties develop their strategies on the basis of their appreciation of the probability that the adjudicator will accept their arguments or the arguments of their adversary. Costs and gains have to be balanced in light of this uncertainty in order to identify the most convenient strategies. This paper provides a probabilistic approach embedded into an argumentation framework to capture this uncertainty and its use to determine the expected utility to engage in a legal dispute.

Keywords. Argumentation, probability, expected utility.

1. Introduction

In legal disputes, there is uncertainty about what statements (the rules and facts asserted by the parties in their arguments) will be accepted by the adjudicator. For the parties involved, uncertainty propagates from the acceptance of their statements to the outcome. In light of this uncertainty expected costs and gains have to be balanced in order to decide whether to start a dispute, and what strategies to adopt in it. This paper provides an approach by which to capture this uncertainty, and it describes the use of this approach to determine the expected utility of engaging in a legal dispute.

In the setting of the argumentation framework by Prakken and Sartor (1997), a probability distribution is assumed with respect to the adjudicator's acceptance of the parties' statements. This distribution results in a probability of an argument's success by taking into account the probabilities of its counterarguments. The former probability is then used in combination with the costs and gains of the dispute in order to predict the expected utility of such a dispute. It is then shown that in deciding whether or not to engage in a legal dispute, the expected utility may provide appropriate guidance to a rational litigant.

¹Supported by the EU projects ONE-LEX (Marie Curie Chair), ESTRELLA (IST-2004-027665), and ALIS (IST-2004-027968).

An important difference with respect to other work in AI on combining logic and probabilities (e.g. Pollock (1995)) is that while the latter considers the probability that a state of affairs holds, we consider the probability that a statement s is accepted by an adjudicator. Such probability can be more precisely expressed as Pr(accepted(q)), which we abbreviate into $Pr(q^a)$. While the laws of probability entail that $Pr(q^a \lor \neg(q^a)) = 1$, they do not entail that $Pr(q^a \vee (\neg q)^a) = 1$. Indeed, there is also a third possibility, namely that neither q nor $\neg q$ is accepted $(\neg q^a \land \neg (\neg q)^a)$. Note that on our interpretation, unlike in the other work, it makes sense to assign probabilities to normative statements or rules. Related work in argumentation on the treatment of uncertainty by the probability of statement acceptance is Roth et al. (2007), but instead of the probability of success of an argument, it proposes to compute the probability that the topic is defeasibly justifiable, where the probability of success of an argument plays no role. By considering the latter probability, the dialogue tree is tracked and that may be a decisive advantage for further investigation in the field of argumentation. Moreover, we want to make as few assumptions as possible on the underlying logic and structure of arguments, to make our ideas as widely applicable as possible.

Let us turn to an example (hereinafter referred to as the mad cow example) to illustrate our approach in this paper. John wants to sue Henry, claiming compensation for the damage that Henry's cow caused to him when he drove off the road to avoid the cow. John knows that he has to show, according to the rule that an animal's owners have to pay damages caused by their animals (a rule the judge will certainly accept), that Henry is the owner of the cow (Henry could claim that the cow belongs to someone else, but since his herd was near the street this is hard to accept) and that the accident was caused by the need to avoid the cow (Henry could claim that the car drove off the road regardless of the cow, but since the cow was hit, this is hard to accept). John knows that even if all of these statements were accepted by the judge, Henry could counterattack in various ways, claiming that the damage was due to John's negligence (according to the rule that when driving on a mountain road one should pay attention to crossing animals, which John did not do to a sufficient degree) or that it was a case of force majeure (the cow suddenly went crazy and crossed into the street). The last objection could be replied to by using the rule (though it is not so sure that the judge will accept it, since it is a debated one) that only exogenous events can count as force majeure, and the cow's madness is endogenous. If John wins, the compensation will amount to 1500, and Henry will also have to pay trial expenses, which are estimated to 500[§]. If John loses he will have to pay trial expenses. For analysing the example we shall introduce atomic sentences that abbreviate the conclusion that is at stake and the reasons that parties adduced in their arguments (see Table 3.1).

So John's problem is the following. According to his analysis of the case, what are the chances which he has of winning? Does it make sense for him to sue Henry?

This paper is organised as follows. In Section 2 we briefly recall the main notions of Prakken and Sartor (1997)'s argumentation logic on which our approach is based. Section 3 provides a method for determining the probability of success of an argument. In Section 4 we briefly address the notion of expected utility. In Section 4, the approach is recapitulated and future investigations are suggested.

2. The argumentation context

Our treatment of uncertainty is embedded in a logical model for argumentation. With regard to the logic supporting the construction of single arguments, we make the following assumptions:

- 1. arguments have a finite nonempty set of premises and a conclusion. No premise of one argument is a conclusion of another argument;
- 2. there is a binary relation of defeat between arguments.

The first assumption is reasonable since if some argument's premise is the conclusion of another argument, the two arguments should ideally be combined by making the second argument a subargument of the first, thus turning the premise of the first into an intermediate conclusion. The second assumptions makes it possible to apply the well-known formal setting of Dung (1995).

We also assume that, given a set of arguments ordered by a binary defeat relation, i.e., given an 'argumentation framework', the dialectical status of arguments is determined with Prakken and Sartor (1997)'s dialectical proof theory for Dung (1995)'s grounded semantics. In this proof theory, a test whether an argument is dialectically justified has the form of a dialogue game between a proponent and an opponent of the argument.

Definition 1 Let T = (Args, defeat) be an argumentation framework. A dialogue on the basis of T is a sequence of moves $move_i = (Player_i, Arg_i)$ $(i > 0, Arg_i \in Args)$, such that

- 1. $Player_i = P$ iff *i* is odd; and $player_i = O$ iff *i* is even;
- 2. *if* $Player_i = Player_j = P$ and $i \neq j$, then $Arg_i \neq Arg_j$;
- 3. *if* $Player_i = P(i > 1)$ *, then* Arg_i *strictly defeats* Arg_{i-1} *;*
- 4. *if* $Player_i = O$ *then* Arg_i *defeats* Arg_{i-1} .

(A strictly defeats B means that A defeats B and B does not defeat A.) The first condition says that player P begins and then the players take turns, while the second condition prevents the proponent from repeating its attacks. The remaining two conditions state the burden of proof for P and O. Figure 1 illustrates two dialogues that can be generated from the mad cow example.

Definition 2 A player wins a dialogue if the other player cannot move.

The idea of this definition is that if P's last argument is undefeated, it reinstates all previous arguments of P that occur in the dialogue, in particular the first one. For example, the proponent does not win the dialogue provided in Figure 2. An argument is shown to be justified if the proponent can make the opponent run out of moves in whatever way the opponent attacks:

Definition 3 An argument A is provably justified on the basis of argumentation framework T iff P has a winning strategy in a game based on T that starts with A.

For example, the argument A of the mad cow example cannot be justified because proponent does not have a winning strategy for A (see Figure 2): if opponent attacks A with B then proponent has no counterattack.

In Prakken and Sartor (1997) this dialectical proof theory was proven to be sound and complete with respect to Dung's grounded semantics. It is easy to see that a winning strategy for A can be displayed as a tree with root A which only branches after P's moves, which contains all possible replies of O and of which all branches end with a move by P. Note also that a strategy for A can be obtained from the game tree for A (i.e., from the tree of all possible games that can be played about A) by deleting at any choice point for P all but one of his possible moves. If these choices can be made in such a way that all remaining branches end with a move by P, the strategy is a winning one for P.

Now as for the adjudicator's acceptances we assume the following setting. When determining whether a claim is sufficiently backed with an argument A, the adjudicator constructs the game tree for A on the basis of a given argumentation framework T and then decides for each premise of each argument in the game tree whether to accept or reject it. Then the adjudicator prunes the game tree by chopping off all branches at the first point where they contain an argument with a rejected premise. Finally, the adjudicator checks in the manner just explained whether the resulting tree contains a winning strategy for proponent. If such a winning strategy exists, this means that A is justified on the basis of a new argumentation framework T' obtained from T by deleting all arguments with a rejected premise and restricting the defeat relation to the remaining arguments.

Now the main problem studied in this paper is how, given a probability distribution over the acceptances of the adjudicator, the probability that argument A is regarded as provably justified in this way by the adjudicator can be determined.

3. Probability of success

In this section we will provide a method for determining the probability of success of an argument by which we mean the probability that the argument is accepted as justified given a knowledge base of which the statements are assigned a probability of acceptance by the adjudicator. Here we shall not discuss the philosophical basis of probability, and in particular the well-known distinction between objective probability and subjective probability, which goes beyond the limits of this paper. We think that our proposal makes sense with regard to both ideas of probability. However, we also believe that the most interesting scenario for the use of our method is for a party's subjective but rational assessment of the chance that his or her argument will stand and overcome all counterarguments given the set of premises which are likely to be accepted by an adjudicator. We recall that the setting we assume is a game tree for an argument A, i.e., the tree of all possible games about A. In this setting, the probability of success of an argument depends upon two conditions:

- the probability that the argument's premises are accepted, which we call *construction chance* (the argument will be rejected if the adjudicator refuses to accept one of the argument's premises);
- 2. the probability that the argument has no valid counterargument, namely no counterargument is able to attack it successfuly, which we call the *security chance* (the argument will be rejected if the adjudicator's acceptances imply that there is a valid attacker of the argument)

In the next sections we discuss how to calculate the construction chance and security chance.

3.1. Construction chance

The construction chance of an argument A, which we denote as Pr(Con(A)) is given by the probability that the adjudicator accepts all premises q_1, \ldots, q_n of the argument, that is, by the probability of $q_1^a, \land \ldots \land q_n^a$, where q_i^a denotes the acceptance of q_i . On the assumption that the premises of an argument are mutually (statistically) independent, Pr(Con(A)) is the product of the probability of acceptance of all premises in the argument:

$$Pr(Con(A)) = Pr(q_1^a \land \ldots \land q_n^a) = Pr(q_1^a) \land \ldots \land Pr(q_n^a)$$
(1)

For instance, the argument A in Figure 1 has three premises namely a, b and the rule r_1 which, respectively, have the probability of acceptance $Pr(a^a) = 1$, $Pr(b^a) = 0, 9$ and $Pr(r_1^a) = 1$. Consequently, the construction chance of A is $Pr(Con(A)) = Pr(a^a).Pr(b^a).Pr(r_1^a) = 0, 9$.



Figure 1. Dialogues generated from the mad cow example.

Literal	Meaning
a	Henry is the owner of the cow.
b	The accident was caused by the need to avoid the cow.
c	Henry has to compensate damages.
d	John was negligent.
e	Cow was mad.
f	Cow's madness is endogenous.

Table 1. Literals with their meaning.

3.2. Security chance

Let us now add the second condition of probability of success of an argument, namely that of not having a successful counterargument. For this purpose we take the view of the justification of arguments as defined in Section 2: recall that all possible arguments (given a set of background information) can be represented as a tree, where each branch of the tree is a dialogue, namely a sequence of arguments by the two parties which defeat one another. For the top argument to be justified, it must be successful along all branches of the tree, which means that it must not be followed by a successful attacker along any dialogue in the tree.

The basic idea is that the chance of success of an argument is diminished to the extent that one of its attackers is going to be successful. Thus, generally, the probability of success of an argument A is diminished by considering the chances of success of all of its counterarguments, along all possible branches of the game tree of which the root is given by A. In the following we shall first consider how to compute the security chance along one dialogue line, and then how to compute security chance along multiple dialogue lines, that is, in a dialogue tree.

3.2.1. Security chance along one dialogue

Let us now precisely characterise the security chance of an argument A_i in a dialogue $D_n = \langle A_1, \ldots, A_n \rangle$, intended as the probability that A_i can really exercise its intended function in the dialogue. Since each argument in a dialogue, according to the dialogue protocol, is defeated (and thus prevented from being successful) by its successor, the probability of success of an argument does not only depend on its construction chance, but also on the chance that the argument's successor fails to be successful. Since both these elements need to be present, we have to deal with the probability of a conjunction of the construction of the argument and the failure of its attacker. We therefore define $Succ(A_i, D_n)$ as $Con(A_i) \land \neg Succ(A_{i+1}, D_n)$. Since equivalent statements are equally probable we have:

$$Pr(Succ(A_i, D_n)) = Pr(Con(A_i) \land \neg Succ(A_{i+1}, D_n))$$
(2)

Applying the product rule $Pr(x \wedge y) = Pr(x) \cdot Pr(y|x)$, the previous result (2) can be rewritten as:

$$Pr(Succ(A_i, D_n)) = Pr(Con(A_i)) \cdot Pr(\neg Succ(A_{i+1}, D_n) | Con(A_i))$$
(3)

Since $Pr(\neg Succ(A_{i+1}, D_n)|Con(A_i)) = 1 - Pr(Succ(A_{i+1}, D_n)|Con(A_i))$, we obtain:

$$Pr(Succ(A_i, D_n)) = Pr(Con(A_i)) \cdot \left[1 - Pr(Succ(A_{i+1}, D_n) | Con(A_i))\right]$$
(4)

We assume that each dialogue implicitly terminates with an empty argument, having no chance of success, so that for the last argument A_n of D_n , the chance of success is given by its chance of construction:

$$Pr(Succ(A_n, D_n)) = Pr(Con(A_n))$$
(5)

For example, the probability of success of argument A along dialogue d2 (see Figure 1) is given by instantiating the equality (4):

$$Pr(Succ(A, D2)) = Pr(Con(A)).[1 - Pr(Succ(C, D2)|Con(A))]$$

= Pr(Con(A)).{1 - Pr(Con(C)|Con(A))
.[1 - Pr((Succ(D, D2)|Con(C))|Con(A))]} (6)

According to the equation (5), the probability of success of D w.r.t. dialogue D2 equals its construction chance, Pr(Succ(D, D2)) = Pr(Con(D)). Furthermore, under the assumption that logically independent statements are also statistically independent, this can be further simplified by eliminating the conditional probabilities (this assumption may not always be warranted). Under these assumptions we obtain:

$$Pr(Succ(A, D2)) = Pr(Con(A)).\{1 - Pr(Con(C)).[1 - Pr(Con(D))]\}\$$

= 0, 9 × {1 - 0, 1 × [1 - 0, 15]} (7)
= 0, 8235

Interestingly, one can demonstrate by induction that the probability $Pr(Succ(A_1, D_n))$ is bounded as follows:

$$Pr(Succ(A_1, D_i)) \ge Pr(Succ(A_1, D_j)) \quad (i \text{ odd}), \tag{8}$$

$$Pr(Succ(A_1, D_i)) \le Pr(Succ(A_1, D_j)) \quad (i \ even), \tag{9}$$

with $j \ge i$. Consequently, the probabilities $Pr(Succ(A_1, D_1))$ and $Pr(Succ(A_1, D_2))$ are respectively the highest bound and lowest bound of $Pr(Succ(A_1, D_j))$. Inside these bounds, the probability $Pr(Succ(A_1, D_j))$ oscillates at every move, with the maximum amplitude that decreases with the length of the dialogue.

3.2.2. Security chance in dialogue tree

We need to consider the probability of success of an argument taking into account that the argument can have more than one counterargument, and that the probabilities of success of such counterarguments somehow need to be added up (since it is sufficient that one counterargument is successful for the argument to fail).

Consider an arbitrary argument A in tree $\tau = \langle \tau_1, \ldots, \tau_k \rangle$ having counterarguments (children) A_1, \ldots, A_k . The probability that any of these counterarguments A_j is successful is the probability of the disjunction $Succ(A_1, \tau_1) \lor \ldots \lor Succ(Arg_k, \tau_k)$, abbreviated as $\bigvee_{j=1}^{j=k} Succ(A_j, \tau_j)$. Then the probability that A is successful is given by the probability that A is constructed times the probability that no such counterargument is successful, that is, that the above disjunction is false:

$$Pr(Succ(A,\tau)) = Pr(Con(A) \land [\neg(\bigvee_{j=1}^{j=k} Succ(A_j,\tau_j))])$$

=
$$Pr(Con(A)).Pr(\neg(\bigvee_{j=1}^{j=k} Succ(A_j,\tau_j))|Con(A))$$
(10)

By transforming the negation of a disjunction into a conjunction of negations ($\neg(x \lor y) = \neg x \land \neg y$), we obtain:

$$Pr(Succ(A,\tau)) = Pr(Con(A)).Pr(\bigwedge_{j=1}^{j=k} \neg Succ(A_j,\tau_j)|Con(A))$$
(11)

To further develop this formula, we need to consider $Pr(\bigwedge_{j=1}^{j=k} \neg Succ(A_j, \tau_j)|Con(A))$ namely the probability that all counterarguments fail to be successful. To do so, we need

to consider an argument's failure under the condition that the other arguments fail as well (since there may be a connection between such failures, as happens when arguments share a common premise). We obtain this result using the idea that $Pr(x \land y)$ is given by Pr(x).Pr(y|x), so that, in general, the probability of a conjunction $a_1 \land a_2 \land a_3 \land \ldots \land a_n$ is given by $Pr(a_1).Pr(a_2|a_1).Pr(a_3|(a_1 \land a_2).\ldots.Pr(a_n|a_1 \land \ldots \land a_{n-1})$. Accordingly, we obtain the following equality:

$$Pr(\bigwedge_{j=1}^{j=k} \neg Succ(A_j, \tau_j) | Con(A))$$

= $\prod_{j=1}^{j=k} \{ Pr(\neg Succ(A_j, \tau_j) | Con(A) \bigwedge_{i=1}^{i=j-1} \neg Succ(A_i, \tau_i)) \}$ (12)

In conclusion by putting together all transformations so far indicated, we obtain:

$$Pr(Succ(A,\tau)) = Pr(Con(A)) \cdot \prod_{j=1}^{j=k} \{ Pr(\neg Succ(A_j,\tau_j) | Con(A) \bigwedge_{i=1}^{i=j-1} \neg Succ(A_i,\tau_i)) \}$$
(13)

To illustrate the above equality, consider Figure 2. The security chance of argument A along the dialogue tree τ is:

$$Pr(Succ(A,\tau)) = Pr(Con(A)).\{Pr(\neg Succ(B,\tau_1)|Con(A) \land \neg Succ(C,\tau_1)) \\ .Pr(\neg Succ(C,\tau_2)|Con(A) \land \neg Succ(B,\tau_2))\}$$
(14)



Figure 2. Dialogue tree generated from the mad cow example.

If we again assume that the premises of the arguments A, B, C, and D are statistically independent, then since trees τ_1 and τ_2 are the dialogues D1 and D2, we can compute (14) as follows.

$$Pr(Succ(A, \tau)) = Pr(Con(A)).Pr(\neg Succ(B, D1)).Pr(\neg Succ(C, D2)) = Pr(Con(A)).[1 - Pr(Succ(B, D1))].[1 - Pr(Succ(C, D2))]$$
(15)
= 0, 9 × [1 - 0, 4] × [1 - 0, 085]
= 0, 4941

Note that the present approach only defines the probability of success of arguments and does not consider the probability whether a statement has at least one successful argument. The difference is especially relevant when a statement can be supported by several arguments. A way to deal with such cases in the present framework is to adopt Prakken (2005)'s treatment of such cases, in which several arguments for the same conclusion can be combined into a new argument for the same conclusion. Then only the probability of success of this single accrual argument needs to be computed. There may be other ways to deal with accrual of arguments but this has to be left for future research.

4. Chances to win and expected utility

We have provided a method to calculate the probability of success of arguments in a dialogue. Although this approach can be applied for different purposes (see Section 5), one of the most direct ways to use it is in calculating the expected utility (EU) of an arguer to take part in the argumentation.

As is well-known, different options are available to calculate EU and there is an extensive debate in the literature. This section is meant to illustrate how to use the method previously developed with the example of the mad cow: complexities about EU are outside the scope of this section. The most classical way to calculate EU of an act X is as follows:

$$EU(X) = \sum_{i=1}^{n} Pr(o_i).u(o_i)$$
(16)

where o_1, \ldots, o_n are the possible (and mutually exclusive) outcomes of X. The equality (16) states that the expected utility of an act X is the sum of the products of the probabilities and utility's values for each outcome, and it can be directly applied to single arguments, thus calculating an arguments' expected utility.

We have two mutually exclusive outcomes: for each arguer, he can win or lose. But we have already defined arguer John as the person who has an interest in starting the argumentation (it is arguer John who wants to sue Henry to obtain compensation from him for damages). Thus, calculating utility mainly makes sense when the act Sue_Henry is considered. If so, only success of argument A corresponds to John's winning, and $EU(Sue_Henry)$ will take into account A's probability of success.

If John wins, the compensation will amount to 1500\$, and Henry will also have to cover the trial expenses (which are estimated to 500\$), so that the compensation will be a net advantage to John. If John loses he will have to cover the trial expenses. Accordingly, the expected utility for John of suing Henry is:

$$EU(Sue_Henry) = Pr(Succ(A,\tau)) \times 1500 + [1 - Pr(Succ(A,\tau))] \times (-500)$$

In the previous section, $Pr(Succ(A, \tau))$ turned out to be 0,4941, and accordingly, $EU(Sue_Henry) = 1500 \times 0,4941 - 500 \times 0,5059 = 488,20$ \$.

The decision to perform an act depends on the expected utility of this act and of its omission. If the expected utility EU(X) of an act X is higher than the expected utility $EU(\neg X)$ of its omission, then a rational agent will perform the act X.

In our example, the expected utility of victory amounts to 488, 20\$ which while the expected utility of not suing Henry is 0\$, since nothing happens in this case (there is no suit and therefore no gain and no loss). Thus, even though *A*'s probability of success is not high, since 488, 20\$ is better then nothing, it makes sense for John to sue Henry.

5. Conclusion

We have provided an approach to capture uncertainty about whether the players' statements will be accepted by an adjudicator and its consequence on the outcome of the dispute. A probability distribution is assumed on these statements to reflect this uncertainty, which induces the probability of success of an argument. The latter probability is then used in combination of the costs and gains of the dispute in order to predict its expected utility. We have illustrated that in deciding in engaging in a legal dispute, the expected utility may be more relevant than the probability of winning the dispute.

An issue is: where do the numbers for the probabilities come from? Statistics could help by providing the number of same-level courts that accepted a specific rule, or by showing how a specific judge decided in the past. If a probability weighting is not empirically verifiable, then a possible refinement is the introduction of variations on it, for example, if John 'feels' that the probability is somewhere around a number X, how much does it matter that John estimates X accurately?

Future works can be generated by variations ranging from the argumentation framework and the uncertainty treatment to the utility function. We intend to couple it with game theory as investigated by Roth et al. (2007) to determine optimal strategies in dialogue games for argumentation where each strategy may have a different utility.

Finally, the approach can be implemented into 'argument assistance systems' that assist the decision making process of jurists and others, by providing a tool that offers advice and reasons for its advice, to engage into a legal dispute.

References

- Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2), 321–358.
- Pollock, J. L. (1995). Cognitive Carpentry: A Blueprint for how to Build a Person. Cambridge, MA, USA: MIT Press.
- Prakken, H. (2005). A study of accrual of arguments, with applications to evidential reasoning. In ICAIL '05: Proceedings of the 10th International Conference on Artificial Intelligence and Law, New York, NY, USA, pp. 85–94. ACM Press.
- Prakken, H. and G. Sartor (1997). Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics* 7(1), 25–75.
- Roth, B., R. Riveret, A. Rotolo, and G. Governatori (2007). Strategic argumentation: a game theoretical investigation. In *ICAIL '07: Proceedings of the 11th International Conference on Artificial Intelligence and Law*, New York, NY, USA, pp. 81–90. ACM Press.