ANGELIC Secrets: Bridging from Factors to Facts in US Trade Secrets

Latifa AL-ABDULKARIM, Katie ATKINSON, Trevor BENCH-CAPON
Department of Computer Science, The University of Liverpool, UK

Abstract. The ANGELIC (ADF for kNowledGe Encapsulation of Legal Information from Cases) project provided a methodology for implementing a system to predict the outcome of legal cases based on a theory of the relevant domain constructed from precedent cases and other sources. The method has been evaluated in several domains, including US Trade Secrets Law. Previous systems in this domain were based on factors, which are either present or absent in a case, and favour one of the parties with the same force for every factor. Evaluations have, however, suggested that the ability to represent different degrees of presence and absence, and different strengths, could improve performance. Here we extend the methodology to allow for different degrees of presence and support, by using dimensions as a bridge between facts and factors. This new program is evaluated using a standard set of test cases.

Keywords. legal case based reasoning, dimensions, factors, facts, ADF.

1. Introduction

The ANGELIC (ADF for kNowledGe Encapsulation of Legal Information from Cases) project [2] provided a methodology for implementing a system to predict the outcome of legal cases based on precedent cases using an Abstract Dialectical Framework (ADF) [7]. In ANGELIC the ADF takes the form of a *tree*, in which the acceptance conditions of the non-leaf nodes are expressed in terms of their children. The higher levels of the nodes represent issues (as in IBP [8]), the leaves represent base level factors and the intermediate nodes represent abstract factors (as in CATO [3]). In [2] the ANGELIC method was evaluated in Trade Secrets Law, the domain of CATO and IBP, and two other domains. The programs produced in [2] treated issues and factors as Booleans. Factors were thus either present or absent in a case, and favoured either the plaintiff or the defendant with every factor having equal force. However, the discussion in [1] suggested representing different degrees of presence and absence and that different extents to which the parties were favoured could improve performance. Here we extend the methodology to allow for different degrees, both of presence and support, by reaching back to facts, and using dimensions [5] as a bridge between facts and factors, as was also done in [11] and [4]. The resulting program has been evaluated using a commonly used set of test cases, namely the 32 cases used in [10].

Table 1. Base Level Factors in CATO

F1	DisclosureInNegotiations	RE-	F15	UniqueProduct	MW+ RE+
F2	BribeEmployee	QM+	F16	InfoReverseEngineerable	LM-
F3	EmployeeSoleDeveloper	LM-	F17	InfoIndependentlyGenerated	LM-
F4	AgreedNotToDisclose	CA+	F18	IdenticalProducts	MW+
F5	AgreementNotSpecific	CA-	F19	NoSecurityMeasures	RE-
F6	SecurityMeasures	RE+	F20	InfoKnownToCompetitors	LM-
F7	BroughtTools	QM+	F21	KnewInfoConfidential	CA+
F8	CompetitiveAdvantage	MW+	F22	InvasiveTechniques	QM+
F10	SecretsDisclosedOutsiders	RE-	F23	WaiverOfConfidentiality	CA-
F11	VerticalKnowledge	LM-	F24	InfoObtainableElsewhere	LM-
F12	OutsiderDisclosuresRestricted	CA+	F25	InfoReverseEngineered	LM-
F13	NoncompetitionAgreement	CA+	F26	Deception	QM+
F14	RestrictedMaterialsUsed	QM+ CA+	F27	DisclosureInPublicForum	LM- RE-

2. Dimensions for CATO

We we will apply our work to the Trade Secrets Domain, as originally analysed in [3], and used in Issue Based Prediction (IBP) [8], AGATHA [9] and previous ANGELIC programs [2]. An ADF based precisely on the abstract factor hierarchy of [3] was given in [2]: we will base our dimensional version on this. The leaf nodes are factors taken from [3]. Our first task is to assign these 26 factors (shown in Table 1) to dimensions. We chose seven dimensions. These include dimensions corresponding to the five values used in [10] (also shown in Table 1), namely confidentiality agreement (CA), questionable means (QM), legitimate means (LM), material worth (MW) and reasonable efforts (RE). Two additional dimensions arise because we distinguish between the physical and contractual security measures taken by the plaintiff and between acquisition by the defendant from practical and library based research. Our seven dimensions are thus:

- The existence of an operational agreement between the parties (Agreement), with value CA.
- The use of illegal or dubious methods by the defendant (Dubious), with value QM.
- The use of legitimate practical means by the defendant (Means), with value LM.
- The security measures taken by the plaintiff (Measures) with value RE.
- The material worth of the information (Worth), with value MW.
- The extent to which the secret had been disclosed by the plaintiff (Disclosure) with value RE.
- The availability of the information (Availability), with value LM.

In Table 2 we show the distribution of CATO's factors to these dimensions. We have given each dimension ten slots, ranging from 10, the extreme pro-plaintiff position, to 0 the extreme pro-defendant position. Factors considered pro-plaintiff in [3] are given slots > 4. The ordering reflects our view of the degree of support, and gaps indicate the size of the range concerned. Note that some factors, such as *RestrictedMaterialsUsed* appear on more than one dimension: that factor can be considered both as an indication of confidentiality, and, since the restrictions were not observed, as the use of dubious means.

Table 2. Dimensions and Their Factors

	Agreement	**Dubious**	**Means**	**Measures**	**Worth**	**Disclosure**	**Availability**
	CA	**QM**	**LM**	**RE**	**MW**	**RE**	**LM**
10	F4	F22		F6	F8		
9		F2					
8	F13	F26	F7		F7	F12	
7	F21	F14					
6	F14				F18		F15
5							
4						F1	
3	F5		F25		F16		F20
2							
1	F23		F17			F10	
0			F3	F19	F11	F27	F24

2.1. Application to ADF

We can now see that the nodes of the ADF are serving several different purposes. This is best seen by looking at the 2-regular ADF of [1]. Some have both children drawn from the same dimension. Both children may favour the same party or they may favour different parties. The former *broadens* the range on the dimension, while for the latter the node allows a choice of which party the dimension favours in the case under consideration. Where both children favour the same party, the abstract factor will be assigned a value corresponding to the maximum value of the factors available for a pro-plaintiff factor and the minimum of the factors available for a pro-defendant factor. Where different parties are favoured a decision is required as to whether the dimension will, as a whole, be taken as pro-plaintiff or pro-defendant. Often it is obvious which should be chosen: for example, if there is a pro-defendant factor on the Agreement dimension, that will be chosen, since it will nullify the pro-plaintiff factors. The factor which most strongly favours the side chosen will determine the value to be passed up.

Other factors relate to two different dimensions. Where the same party is favoured on both dimensions, the most favourable factor to the favoured party provides the value: where the points on the dimensions conflict it must first be determined which party the abstract factor will favour. Thus the plaintiff may be favoured on the Disclosure dimension because the disclosures were restricted (F12) and the defendant may be favoured on the Means dimension because they had reverse engineered the information (F25). In such a case the court will need to decide the relative part played by the reverse engineering and the restricted disclosure in the defendant's acquisition of the knowledge. We will resolve this according to our view of value preferences. We will take the value of the abstract factor from the dimension with the preferred value, although some moderation of the effect of the chosen dimension could be applied if desired.

2.2. Relation to Values and to ANGELIC CATO

In [10], an empirical evaluation was presented of the proposals of [6]. The version in [10] using dimensions based them on values. Five values were used in [10] and

these same five values underpin our dimensions. We have however in the case of Legitimate Means and Reasonable Efforts distinguished between physical methods and research-based methods: therefore we have two dimensions promoting Legitimate Means (means and availability) and two for Reasonable Efforts (measures and disclosure). Thus the primary role of values is to underpin the dimensions. Preferences are used when we need to balance dimensions, and revealed by the way courts make their choices.

The ANGELIC project was designed to follow the CATO analysis faithfully. Thus the original program took the structure of the factor hierarchy in [3] and used it directly (with an issue level from [8]) to provide the nodes for the ANGELIC ADF. This was further developed as 2-regular ADF in [1]. The use of dimensions allows us to simplify this structure. Where the children of a node are both from the same side of the same dimension, they can be replaced by the point most favourable to the relevant party. Where they are from different ends of the same dimension, a choice must be made. If we make this choice in general at the outset, we will require abstract factors only to merge two dimensions, or resolve a conflict between them.

3. Evaluation

We use the 32 cases used in [10] (these are all the cases of the 180 used [8] that are available in the public domain). To realise a program from the ADF we first represented the cases by providing a number for the best plaintiff point and best defendant point on each dimension for each case. Where there were no points on a dimension in a case we used a default value. The party favoured by the defaults reflected the practice in the cases: while the plaintiff needs to provide evidence of dubious behaviour, the defendant must provide evidence with respect to the other issues. Thus the plaintiff is favoured by the default, except on the Dubious Methods dimension. Where there are two different values on a dimension these are reconciled, as described in section 2.1 to produce a single value for each dimension. These are then propagated through nodes which merge and choose between dimensions, again as described in section 2, until the issue layer is reached. At this point the value of the abstract factor is compared with a threshold (set that it is just above or below the default as appropriate), and then the logical expression in terms of the issues can be evaluated.

Of the thirty-two cases, twenty-nine were decided correctly and three, Goldberg, Space Aero and Scientology were decided incorrectly. Comparison with previously reported experiments is shown in Table 3.

Note that ANGELIC Secrets performs better than ANGELIC CATO: this appears to be because the use of dimensions has provided a way of countering the effect of F16 (ReverseEngineerable), which caused difficulties with the earlier ANGELIC program. In Goldberg the discussion in [2] suggested that F27 should not be considered present since the decision states:

> The district court found that Medtronic could not avoid its obligation of confidence due to the availability of lawful means of obtaining the concept when those means were not employed. We affirm.

Table 3. Results, including some from [8], [9] and [2]

	correct	error	abstain	accuracy	no-abst
ANGELIC Secrets II	31	1	0	96.8	96.8
ANGELIC Refined	31	1	0	96.8	96.8
AGATHA Brute Force	30	2	0	93.7	93.7
IBP	170	15	1	91.4	91.9
AGATHA A*	29	3	0	90.6	90.6
ANGELIC Secrets	29	3	0	90.6	90.6
CATO-coin	163	30	0	89.0	89.0
Naive Bayes	161	25	0	86.5	86.5
HYPO-coin	152	34	0	81.7	81.7
ANGELIC CATO	25	7	0	78.1	78.1
CATO	152	19	22	77.8	88.8
HYPO	127	9	50	68.3	93.4
IBP-model	99	15	38	72.6	86.8

Changing the point on the disclosure dimension to the default gives the correct decision. In [2] Space Aero gave the plaintiff problems arising from the lack of security measures (F19) and the absence of any factors to establish a confidential relationship. ANGELIC Secrets also fails on these two issues. But the decision in Space Aero explicitly states that the security measures were adequate:

> The testimony, taken as a whole, convinces us that Darling took precautions to guard the secrecy of its process which, under the circumstances, were reasonably sufficient.

so, at worst, the default value should be used rather than F19. With respect to confidentiality, the decision was based on the fact that the defendants were former employees and had acquired the information in that capacity, and that this imposed a duty of fidelity sufficient to establish a confidential relationship. The ex-employee problem arises in several cases, so perhaps we should introduce an additional dimension point to represent this fact. Meanwhile if we replace F19 with the default on the Measures dimension, and we add F21 to indicate that the defendants were aware that the information was confidential, ANGELIC Secrets makes the correct prediction. The modified version is shown in Table 3 as ANGELIC SECRETS II.

The misclassified case in ANGELIC SECRETS II is *Scientology*, which was not analysed in [3] and appears for the first time in [8]. The *Scientology* case has factors F4, F6, F10, F12 and F20, and given these factors it is hard to see how the plaintiff lost, since any disclosures appear to have been covered by agreements not to disclose. But another aspect seems to have come into play here. As reported in the *LA Times* of 1986-08-09:

> In its unanimous opinion, a three-judge panel of the U.S. 9th Circuit Court of Appeals said the state law protects economic, but not religious, secrets.

Thus failure on this case may result from the case falling outside the domain, so that other issues would come into play. We will therefore not suggest refinements to accommodate this case. Excluding *Scientology* would give 100% success.

The evaluation has shown that using dimensions to bridge to facts can provide some improvement to the performance of the system. In particular, when compared with the same case representation previously used in [2], five of the seven cases wrongly classified there are now decided correctly, because of a better handling of reverse engineerability in the dimensions set up. Two cases are misclassified in both ANGELIC CATO and ANGELIC Secrets, but these can be explained by referring to the texts of the decisions. One further case is misclassified, but there is evidence to suggest it does not really form part of the domain at all.

4. Concluding Remarks

In this paper we have shown how dimensions can be used to bridge between factors and case facts, building from the ANGELIC Methodology presented in [2]. Not only does this provide a useful extension in that it allows the attribution of factors (dimension points) to be made explicit, and varying degrees of support to be accommodated, but performance is improved. Moreover the use of dimensions allows for expression of the cases in a narrative form, as in [4]. We consider that this contributes a vital component to the ANGELIC methodology.

References

[1] L. Al-Abdulkarim, K. Atkinson, and T. Bench-Capon. Factors, issues and values: Revisiting reasoning with cases. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law*, pages 3–12, 2015.

[2] L. Al-Abdulkarim, K. Atkinson, and T. Bench-Capon. A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence and Law*, 24(1):1–49, 2016.

[3] V. Aleven. *Teaching case-based argumentation through a model and examples*. PhD thesis, University of Pittsburgh, 1997.

[4] T Bench-Capon and F Bex. Cases and stories, dimensions and scripts. In *Proceedings of JURIX 2015*, volume 279, pages 11–20. IOS Press, 2015.

[5] T. Bench-Capon and E. Rissland. Back to the future: dimensions revisited. In *Proceedings of JURIX 2001*, pages 41–52. IOS Press, 2001.

[6] T Bench-Capon and G Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1):97–143, 2003.

[7] G Brewka, S Ellmauthaler, H Strass, J Peter Wallner, and S Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the 23rd IJCAI*, pages 803–809. AAAI Press, 2013.

[8] S. Brüninghaus and K. Ashley. Predicting outcomes of case-based legal arguments. In *9th International Conference on Artificial Intelligence and Law*, pages 233–242, 2003.

[9] A Chorley and T Bench-Capon. Agatha: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, 13(1):9–51, 2005.

[10] A. Chorley and T. Bench-Capon. An empirical investigation of reasoning with legal cases through theory construction and application. *Artif. Intell. Law*, 13(3-4):323–371, 2005.

[11] H Prakken, A Wyner, T Bench-Capon, and K Atkinson. A formalization of argumentation schemes for legal case-based reasoning in aspic+. *Journal of Logic and Computation*, 25(5):1141–1166, 2015.