A Logic-based Architecture for Opinion Dynamics

Piter Dykstra[12], Corinna Elsenbroich[3], Wander Jager[1], Gerard Renardel de Lavalette[1], Rineke Verbrugge[1]

[1] Groningen University, The Netherlands P.Dykstra at rug.nl
[2] Hanzehogeschool Groningen
[3] University of Surrey, Guildford, United Kingdom

Abstract. We explore the usefulness of Fuzzy Logic for the implementation of dialogical logic into an agent-based model of opinion extremisation. We tackle a pervasive problem in argumentation logics, that linguistic incompatibility is not the same as logical inconsistency. Using Fuzzy Logic, the linear ordering of opinions is replaced by a set of partial orderings leading to a new, extra-logical property/attribute/notion of consistency in the form of *convexity* of sets. By doing so we also get close to an implementation of concepts of the Theory of Reasoned Action, i.e. the interplay between the importance and the evidence of a belief, and of Social Judgement Theory such as latitudes of acceptance, rejection and non-commitment. The fuzzy extension not only allows for modelling the emergence of processes of e.g. polarisation, pluriformity and other aggregated phenomena of social interaction on opinions, but it also allows for studying the downward causation process in more detail.

Introduction

Social systems involve the interaction between large numbers of people, and many of these interactions relate to the shaping and changing of opinions. The dynamics of these opinions elucidate other relevant phenomena such as social polarisation, extremism and multiformity. It is, however, important to develop models that capture interaction processes in a simple yet realistic manner. We want to model *discussions* between agents and we consider a number of concepts as critical: certainty of beliefs (*evidence*), the *tolerance* level for (partly) deviating opinions, the *importance* of a belief and agent's *reputation*. In this paper we combine principles of Social Judgment Theory [11] with principles of Fuzzy Logic [12], [7]. Using the Fuzzy Logic approach allows for a more realistic modelling of the gradient between assimilating another agent's position, responding neutrally towards this position, or contrasting one's own position.

The Theory of Reasoned Action [1] defines an attitude as the sum of beliefs about a particular behaviour weighted by the evaluations of these beliefs. In our model the attitude is a sum of opinions, each having an own importance. Since beliefs are evaluative by nature, they can be interpreted as opinions as well. And because the evaluation of beliefs refers to their relative importance, our conceptual modeling is also in accordance with the main principles of the Theory of Reasoned Action.

Social Judgement Theory [11] describes the conditions under which a change of attitudes takes place, with the intention to predict the direction and extent of that change. Jager and Amblard [10] demonstrate that the attitude structure of agents determines the occurrence of *assimilation* and *contrast* effects, which in turn cause a group of agents to reach consensus, to bipolarize, or to develop a number of subgroups sharing the same position. In this theory agents engage in social processes. However their framework does not have a logic for reasoning about the different agents and opinions.

In earlier work [6], we designed a framework for the simulation of public opinion dynamics based on the idea that the dynamics of social relations is interrelated with the dynamics of opinions. This model was designed to fit an agent communication model based on dialogues. In contrast to traditional opinion dynamics models [9], [8], [10] we developed a model in which agent communication is based on dialogues and an agent's social group. We follow Carley's [3, 4] social construction of knowledge, meaning knowledge is determined by communication dependent on an agent's social group while the social group is itself determined by the agents' knowledge. Elements of our baseline model as presented in [6] are the following[4]:

- *Argumentation* — opinions are expressed as arguments, i.e. linguistic entities defined over a set of literals rather than points on a line.
- *Game structure* — agents play an argumentation game in order to convince other agents of their opinions. The outcome of a game is decided by vote of the surrounding agents.
- *Reputation status* — winning the dialogue game results in winning reputation status points.
- *Social network* — who plays with whom is restricted by the mutual distance.
- *Alignment of opinions* — agents adopt opinions from other agents in their social network/social circle (on the space).

In addition, the extended model presented in this paper has:

- *Attitudes* — The change in belief in individual statements forces the reevaluation of the agent's attitude.
- *Partial Orderings* — Normally atomic statements are considered as an unordered set, and in Social Judgement Theory there is a continuous ordering on statements. We assume a set of partial ordering on literals, one for each issue. A literal is an atomic statement or its negation. An issue is a scale by which statements may be compared with each other as being closer to the ideal of that issue (e.g. health awareness or political orientation). By $S_1 \leq_I S_2$ we mean that statement S_2 is closer to some positive ideal of issue I than S_1 is. For the sake of simplicity, we assume that the ordering on literals is the same for all agents.
- *Linguistic consistency* — Linguistic consistency and logical consistency are not the same. A logically consistent agent cannot believe a statement and its negation, but in order to be linguistically consistent, an agent A_1, who

[4] An implementation in Netlogo of the baseline model is available at: *http://www.math.rug.nl/ piter/essa/index.html*

believes S_1 and S_3, also has to believe all the positions that are intermediate between A and C. So when $S_1 \leq_I S_2 \leq_I S_3$ holds for any issue I, A_1 also has to believe S_2.

The paper is structured as follows. In Section 1 we briefly introduce the baseline model built on dialogical logic. In Section 2 we extend the model with fuzzy logic by adding an ordering on atomic statements [2], [11]. In Section 3 we conclude and point to future work.

1 A Model with 2-dimensional Belief Values

Our framework for agent knowledge developed in [6] was a multi-valued extension of Belnap & Anderson's 4-valued logic [5], designed to cope with situations in which agents get different degrees of evidence, possibly contradictory, about a statement.

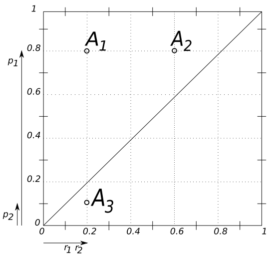
Agents are located in a space where they form groups. They try to stay in tune with the opinions of their surrounding agents. Agents assume that the common opinion on a certain location is the accumulation of all utterances in that area that are not defeated in a debate. The real common opinion, however, is determined by the vote of the neighbourhood agents about atomic statements at the end of a debate. Agents want to gather reputation points, which is only possible by winning debates. They like to attack utterances of other agents if they think they can win the debate. They also like to make novel, previously unheard, utterances in order to align common opinion with their own opinion.

The argumentation game is played as follows: An announcement consists of a statement S and two numbers p, r between 0 and 1. When an agent announces (S, p, r), it says: 'I believe in S and I want to argue about it: if I lose the debate I pay p reputation points, if I win I receive r reputation points'. The values p and r are designed to be used in the communication game, but not very useful when we want to answer the question: *"What will an agent learn from an announcement made by another agent?"* To answer that question we introduce two more intrinsic concepts: the *degree of belief or the evidence* (e) an agent has in favour of a statement and the *importance* (i) of that statement.

Evidence is the ratio between pro and con. An evidence value of 1 means *absolutely convinced of S*. An evidence value of 0 means *absolutely convinced of the opposite of S*. *No clue at all* is represented by 0.5. An importance value of 0 means *unimportant* and 1 means *most important*. Importance is expressed by the sum of both odds. This means the more important an issue the higher the odds:

$$e = \frac{p}{p+r} \qquad i = \frac{p+r}{2}$$

The (p, r)-values can also be computed from an (e, i)-pair: $p = 2 \cdot e \cdot i$ and $r = 2 \cdot i \cdot (1 - e)$. The two rightmost columns in the table of Figure 1 show the (e, i)-values corresponding to the (p, r)-values. We adopt an *accept*-operation, which, when applied to two belief values, (e_1, i_1) and (e_2, i_2), of a propostions S, will produce a belief value that is acceptable for an agent hearing S with



Agent	*Pay*	*Reward*	*Evidence*	*Importance*
A_1	0.8	0.2	0.8	0.5
A_2	0.8	0.6	0.571	0.7
A_3	0.1	0.2	0.333	0.15

Fig. 1. The influence of communication on common opinion. The points A_1, A_2, A_3 in the square of possible belief values represent the odds that three agents, A_1, A_2 and A_3, have assigned to a certain statement, S. Let agent A_1 announce that it wants to defend statement S against any opponent who is prepared to pay $r_1 = 0.2$ reputation points in case A_1 wins the debate. A_1 declares itself prepared to pay $p_1 = 0.8$ points in case it loses the debate. A_2 agrees with A_1 (they are both in the N-W quadrant), but the two agents have different valuations. A_2 believes that the reward should be bigger, $r_2 = 0.6$, on winning the debate, meaning A_2 is less convinced of S than A_1. A_3 disagrees completely with both A_1 and A_2. A_3 believes that the reward should be twice as high than the pay ($p_3 = 0.1$ against $r_3 = 0.2$). This means A_3 believes it is advantageous to attack A_1, because it expects to win 0.8 points in two out of three times, with the risk of having to pay 0.1 points in one out of three times.

(e_1, i_1) while it believes S with (e_2, i_2). The *accept*-operation should implement the following behaviour. If an agent is in agreement with a proposition S, repeated utterances of the same message count as evidence, if in disagreement, repeated utterances detract evidence in favour of S. In order to preserve evidence and beliefs agents can ignore utterances or move away from the place of utterance.

We need two operators, $\Uparrow$ $\Downarrow$ for growing and shrinking fulfilling the following constraints: (1) Boundedness: $0 < a \Uparrow b < 1$, $0 < a \Downarrow b < 1$. (2) Monotonicity: $a, b < a \Uparrow b$ for asymptotically growing and $a \Downarrow b < a, b$ for shrinking. (3) $\Downarrow$ is related to the harmonic mean (cf. Definition 1) (4) $\Uparrow$ is a dual form of shrinking. (cf. Definition 2)

Definition 1 (Shrink).

$$a \Downarrow b =_{def} \frac{a \cdot b}{a + b}$$

Definition 2 (Grow).

$$a \Uparrow b =_{def} 1 - (1 - a) \Downarrow (1 - b)$$

Both operators are *commutative* and *associative* but not *idempotent*. *Grow* and *shrink* work well for the description of the behaviour of *importance* but not for *evidence*. Growing evidence boils down to moving away from 0.5 to 0 or 1, whichever is closest. The shrinking of evidence, which is only meaningful in the case of opposite opinions, is expressed by moving towards 0.5, which can be implemented by the taking arithmetic mean.

Definition 3 (Growing Evidence).

$$a \Uparrow_E b =_{def} \begin{cases} a \Uparrow b & \textbf{\textit{if}} \quad a \geq 0.5 \wedge b \geq 0.5 \\ a \Downarrow b & \textbf{\textit{otherwise}} \end{cases}$$

Definition 4 (Shrinking Evidence).

$$a \Downarrow_E b =_{def} (a+b)/2$$

Now we can define validation and invalidation in terms of growing and shrinking of the belief components *importance* and *evidence*.

Definition 5 (Validation).

$$(e_1, i_1) \oplus (e_2, i_2) =_{def} \quad (e_1 \Uparrow_E e_2, \quad i_1 \Downarrow i_2)$$

Definition 6 (Invalidation).

$$(e_1, i_1) \ominus (e_2, i_2) =_{def} \quad (e_1 \Downarrow_E e_2, \quad i_1 \Uparrow i_2)$$

With the help of the previous definitions we can define what we mean by "acceptance of an announcement S, with belief value (e_1, i_1) by an agent who believes statement S with value (e_2, i_2)". By $(e_1, i_1) Ⓐ (e_2, i_2)$ we mean the belief value the agent has w.r.t. S after accepting that announcement.

Definition 7 (Acceptance).

$$(e_1, i_1) Ⓐ (e_2, i_2) =_{def} \begin{cases} (e_1, i_1) \oplus (e_2, i_2) & \textbf{\textit{if}} \quad (e_1, e_2 \geq 0.5) \vee (e_1, e_2 < 0.5) \\ (e_1, i_1) \ominus (e_2, i_2) & \textbf{\textit{otherwise}} \end{cases}$$

2 Attitudes: Aggregated Statements

So far we considered only one statement in our model of opinion dynamics. We did not account for a meaningful relation between different statements, e.g. in the sense that S_1 is more extreme than S_2 on some extremeness-scale. Thus we could not describe the change of the state of belief to a more (or less) extreme position.

An agent's attitude can be understood in terms of what statements it finds acceptable or not. These sets of opinions are referred to as the latitudes of *acceptance* (LOA), *rejection* (LOR), and *non-commitment* (LON). We integrate these latitudes in our architecture as *fuzzy sets* [12, 13],[7], because agents already use *degrees of belief*.

The LOA is determined by an interval of a certain size around a *preferred opinion*. This interval is surrounded by the latitudes of non-commitment, which border on their other side to the area with the statements that are rejected. We consider latitudes as discrete sets, since they represent statements. However, we allow statements to be *partially ordered* in an extremeness scale according to some *issues*.

The LOA is *not interrupted*. If an agent accepts two statements with different positions on the scale of extremeness, it also has to accept all intermediate statements. This principle will be called: *linguistic consistency*.

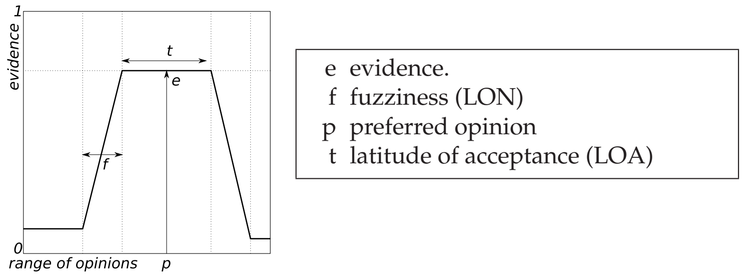


Fig. 2. The membership function of a fuzzy set. In Fuzzy Logic the membership of a fuzzy set S is described by a function, $\alpha_S : dom(X) \rightarrow [0,1]$, where $dom(X)$ is the domain of the *linguistic variable* X, (e.g. temperature is a linguistic variable for fuzzy predicates as cold, moderate, warm, etc). In our case: $dom(X)$ is the set of statements in the model, $Props$, and for the range $[0,1]$ we take the set of belief values $Belief$-$values$. We let the linguistic variable denote a statement and the outcome of the membership function is the belief value an agent assigns to that statement. The LOA is defined as the set of all statements with an evidence value higher than a certain e_2-value; the most preferred opinion is one of the maximal $e-$values of that set. The LOR is the set of statements with evidence values lower than a certain e_1-value, with $0 \leq e_1 < e_2 \leq 1$. Propositions with $E-$values greater than e_1 and less or equal than e_2 form a set representing the LON.

Definition 8 (Attitude). *An attitude is a function* $\alpha : Props \rightarrow Belief\text{-}values$

Typical (fuzzy-) set operations are *union* and *intersection*. They can be defined in terms of operations on their elements, e.g. *union* in terms of *max*: $(\alpha_A \cup \alpha_B) =_{def} max(\alpha_A(x), \alpha_B(x))$. In the same way we can extend our concepts for belief acceptance to attitudes. Validation can be defined as:

$$(\alpha_A \oplus \alpha_B)(x) =_{def} \alpha_A(x) \oplus \alpha_B(x)$$

and similarly for the definitions of invalidation and acceptance. Our attitudes need to be proper intervals (latitudes):

Definition 9 (Convexity). *An attitude* $\alpha : dom(X) \rightarrow Belief\text{-}values$ *is convex when:*

$$\forall x, y, z \in Props : ((y \leq x \leq z) \rightarrow min(\alpha(y), \alpha(z)) \leq \alpha(x))$$

or more specific: $\alpha(x) = min(sup\{\alpha(y) \mid y \leq x\},\ sup\{\alpha(z) \mid x \leq z\}) \quad (*)$

It is possible to make an attitude convex by increasing the belief values of the statements that violate the convexity-constraint sufficiently:

Definition 10 (Convexification). *The convexification,* α^c*, of an attitude,* α *is:*

$$\alpha^c(x) =_{def} \ min(sup\{\alpha(y) \mid y \leq x\},\ sup\{\alpha(z) \mid x \leq z\})$$

From $(*)$ it follows that: α^c is convex; $\alpha \leq \alpha^c$; if α is convex than $\alpha^c = \alpha$; from α_1 is convex and $\alpha \leq \alpha_1$ follows $\alpha^c \leq \alpha_1$.

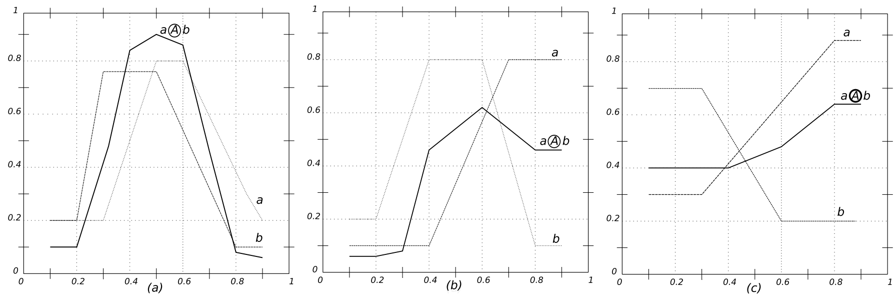


Fig. 3. The change of attitudes as a result of accepting other attitudes. (a) Similar attitudes enlarge certainty. (b) Two related attitudes result in one with a greater tolerance. (c) Unrelated attitudes result in one with less certainty and less tolerance.

The examples in fig. 3 we can see how the acceptance operator on complete attitudes produces behaviour as predicted in Social Judgement Theory and the Theory of Reasoned Action. Fig. 4 shows that the choices of changes in attitude of agents are nontrivially constrained by convexity.

3 Conclusion

We have extended our dialogical logic based simulation architecture with Fuzzy Logic to represent agent attitudes. We introduced a set of a partial orderings on statements, which serves as the domain of the linguistic variables for the corresponding fuzzy sets. The fuzzy set membership function describes the degree of belief in a preferred opinion and the tolerance level (latitude) on which agents decide on the truth of statements. To capture the notion of latitudes in Social Judgement Theory we introduced the notion of *linguistic consistency*, which constrains the agent's reasoning. With this extension, changes of opinions can be expressed according to the Social Judgement Theory. This enterprise is intended to simulate processes of extremisation in society. We have shown that the latitudes of acceptance of individual agents can be adjusted by the outcome of their interactions with other agents in dialogue games (downward causation). This architecture allows the social opinion dynamics according to the Theory of Reasoned Action in simulations with reasoning and debating agents to emerge. Future work is the implementation of a fuzzy reasoning engine into the simulation framework.

References

1. I. Ajzen and M. Fishbein. *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall, 1st edition, 1980.
2. K. J. Arrow. *Social Choice and Individual Value*. John Wiley and Sons, New York, 2nd edition, 1963.

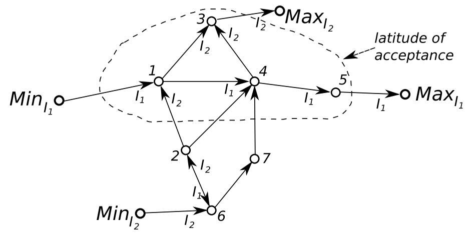


Fig. 4. A 2-dimensional convex LOA. The points represent statements and the *Min*- and *Max*-points represent extreme positions, with respect to two issues, I_1 and I_2. The arrows are annotated with an I_1 or an I_2 if the propositions are only related with respect to the I_1 or I_2 issue respectively. If there is no annotation the arrow represents both orders. The area with the dashed border represents the latitude of acceptance of an agent. There are four statements accepted by the agent. It may drop a statement, except proposition 4, because that would violate the convexity-principle. The agent may extend its latitude in four directions (or any combination of those directions). It cannot add proposition 2 without adding proposition 6 and 7, nor can it add proposition 7 without adding proposition 2 and 6.

3. K. M. Carley. Knowledge acquisition as a social phenomenon. *Instructional Science, Elsevier Science Publishers B.V., Amsterdam*, 14(4):381–438, 1986.
4. K.M. Carley, J. Reminga, and N. Kamneva. Destabilizing terrorist networks. In *NAACSOS conference proceedings*, 2003.
5. Belnap, Jr. N. D. A useful four-valued logic. In J. Michael Dunn and George Epstein, editors, *Modern Uses of Multiple-Valued Logic*, pages 5–37. D. Reidel Publishing Co., Dordrecht, 1977.
6. P. Dykstra, C. Elsenbroich, W. Jager, G. Renardel de Lavalette, and R. Verbrugge. A dialogical logic-based simulation architecture for social agents and the emergence of extremist behaviour. In B. Edmonds and N. Gilbert, editors, *The 6th Conference of the European Social Simulation Association*, page 37. Electronic Proceedings 1844690172, 2009.
7. P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.
8. Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
9. S. Huet, G. Deffuant, and W. Jager. A rejection mechanism in 2d bounded confidence provides more conformity. *Advances in Complex Systems (ACS)*, 11(04):529–549, 2008.
10. W. Jager and F. Amblard. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational & Mathematical Organization Theory*, 10(4):295–303(9), 2005.
11. M. Sherif and C.I.Hovland. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change.* New Haven: Yale University Press., 1961.
12. L.A. Zadeh. Fuzzy sets. *Information & Control*, 8:338–353, 1965.
13. L.A Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 1:119–249, 1975.