Uncertainty and fuzziness: from natural language to argumentation models

Pietro Baroni
DII - Dip. di Ingegneria dell'Informazione
University of Brescia (Italy)
Based on joint work with
Massimiliano Giacomin and Beishui Liao

Uncertain and fuzzy statements!

- I believe that tomorrow will probably be a bit colder than today because it seems that a northern strong wind is coming

- Explicit uncertainty and fuzziness are present in many natural language statements

Uncertain and fuzzy statements?

- Tomorrow will be rainy
  because the weather forecast says so

- Implicit uncertainty and fuzziness pervade most (almost all?) natural language statements

Arguments in natural language

- Humans do argue!
- Argumentative structures too pervade most (~~almost~~ all?) natural language statements
- Argument mining aims at "automatically identifying argumentative structures within a document, e.g., the *premises, conclusion, and argumentation scheme* of each argument, as well as *argument-subargument and argument-counterargument relationships* between pairs of arguments in the document”

Argumentation and uncertainty / fuzziness

- They are intimately close, even compenetrating, in daily life …
- … but they look like living a sort of "in-house separation" if one considers the relevant formal or semi-formal models
- Can we hope in a happy marriage in the end?

Formal argumentation models

- Abstract argumentation formalisms
  - » Dung's AFs
  - » Brewka-Woltran's ADFs
  - » Variants of AFs (bipolar, value-based, preference-based)
- Focused on arguments as abstract entities (their structure, if any, is ignored) and on their relationships (traditionally mainly attacks, but also support, …)
- Steps far from natural arguments

Formal argumentation models

- Semi-abstract structured argumentation formalisms
  - » ASPIC+
  - » Vreeswjik AASs
  - » ABA frameworks
- Formalisms dealing with the structure of arguments and their relationships in a language-independent way
- Closer to natural argument structure but still abstract

Formal argumentation models

- Fully instantiated argumentation formalisms
  - » DeLP
  - » Classical Logic Based
  - » TOAST-ASPIC
- Fully implementable (and implemented) treatment of argument construction, interaction and evaluation (possibly based on more abstract formalisms in some parts)
- Fully equipped to represent natural arguments, in principle, but typically have "unnatural" roots

Semi-formal argumentation models

- Argument schemes are a well–known informal but structured approach to analyze and characterize arguments
- Argument schemes use structured natural language descriptions
- Argument schemes have been used in many applications as a first modeling tool to capture argumentation occurring "into the wild"
- Identification of argument schemes is a key element of argument mining in natural language

Uncertain arguments

- Argumentation is "uncertain by nature":
  - » Arguments are generally regarded as defeasible
  - » Multiple alternative evaluation results (extensions/labellings) are available
  - » Dynamic process subject to unforeseeable evolutions
- Purely symbolic/qualitative/crisp uncertainty
- No degrees (sometimes a partial preference order that might correspond to uncertainty degrees)
- Far from the expressivity needs of natural language

Fuzzy arguments

- Is argumentation "fuzzy by nature?"
- If it is, traditional argumentation models ignore it
- They are all based on crisp sets
- No fuzziness at all

- Far from the expressivity needs of natural language

Bringing uncertainty and fuzziness to argument models!

- Probabilistic argumentation has become a "hot topic" within the community in recent years
- Some fuzzy argumentation has appeared too

- The common underlying idea is that formal argumentation needs to import some "additional features" from other research areas that are:
  - older
  - more developed
  - more basic

Bringing argumentation to uncertainty and fuzziness models?

- Is argumentation as a whole:
  - » younger
  - » less developed
  - » less basic

  than probability theory or fuzzy set theory?

- Is the idea of exploring an argumentation-based interpretation of probability or of fuzziness farfetched?

Uncertainty and fuzziness embeddings

- Uncertainty (mainly, but not only, probabilistic) has been embedded in both abstract (a lot of works), semi-abstract (one work) and fully instantiated (several works) formalisms
- Fuzziness too has been considered in abstract, semi-abstract, and fully instantiated argumentation formalisms (a few works each)
- Interesting approaches but uncertainty/fuzziness modeling seems to occur "too late" or at "safety distance" from natural uncertainty/fuzziness sources

Reverse engineering

- One could proceed top-down from uncertainty/fuzziness modelling at the more abstract levels towards uncertainty/fuzziness modelling at the more concrete levels
- A sort of reverse engineering (abstract models looking for applications):
  - » What the probability attached to this attack relation might mean?
  - » What the probability attached to this ASPIC rule might mean?

Reversing reverse engineering

- A top-down perspective is easier to start with, is intellectually stimulating and can shed some light on otherwise unattackable matters

- But …

  is it natural?

- And …

  will it work in the end?

The missing link

- A bottom-up perspective from application requirements to model definition appears at least as worth exploring as the top-down one

- This points to a missing link: dealing with uncertainty and fuzziness in argument schemes

Argument schemes

- Semiformal model, using structured natural language descriptions
- Premises (sometimes accompanied by qualifications like Major, Minor ...)
- A “stereotypical” reasoning pattern (synthesised by the scheme name) connecting the premises to a defeasible conclusion
- Some critical questions pointing out potential weaknesses to be identified within instances of the scheme

Argument schemes

- Argument schemes are a very flexible and intuitively appealing approach to start modeling arguments on field
- Direct relations with common-sense examples
- Sixty primary schemes (many with subschemes) in the Walton-Reed-Macagno 2008 book, many adaptations and variations in specific papers
- Approximately half of the applications presented at the COMMA conference series use argument schemes (often in combination with other formalisms)

Scheme mining

- Argument schemes are an ideal first target for argument mining activity
- Even just trying to identify the premises and the conclusion corresponds to look for a (very simple) argument scheme
- But …

Too open and flexible?

- Argument schemes per se are a rather “elusive” approach just because they are so open and flexible
- One size fits all but …
- is it really one size or just stretched every time according to the needs? (changing size fits all)
- Argument schemes are really heterogeneous: different schemes seem to rely on different modelling assumptions and constraints
- The borderline between what is included explicitly and what is left implicit is fluid

Implicit and explicit linguistic uncertainty

- Scheme APK (Argument from position to know)

Major Premise: Source a is in a position to know about things in a certain subject domain S containing proposition A.

Minor Premise: a asserts that A (in domain S) is true (false).

Conclusion: A is true (false).

CQ1: Is a in a position to know whether A is true (false)?

CQ2: Is a an honest (trustworthy, reliable) source?

CQ3: Did a assert that A is true?

- APK has no elements of explicit uncertainty inside
- CQs represent some possible doubts
- Some terms in the CQs are fuzzy (e.g. honest)

Implicit and explicit linguistic uncertainty

- Scheme ACE (Argument from cause to effect)

Major Premise: Generally, if A occurs, then B will (might) occur.

Minor Premise: In this case, A occurs (might occur)

Conclusion: Therefore, in this case, B will (might) occur.

CQ1: How strong is the causal generalization?

CQ2: Is the evidence cited (if there is any) strong enough to warrant the casual generalization?

CQ3: Are there other causal factors that could interfere with the production of the effect in the given case?

- ACE has elements of explicit uncertainty inside
- Some terms in the CQs are fuzzy (e.g. strong)

Where is the difference?

- APK with uncertainty

Major Premise: Source a is (possibly) in a position to know about things in a certain subject domain S containing proposition A.

Minor Premise: a asserts that A (in domain S) is (might be) true (false).

Conclusion: A is (might be) true (false).

- ACE without uncertainty

Major Premise: Generally, if A occurs, then B will (might) occur.

Minor Premise: In this case, A occurs (might occur).

Conclusion: Therefore, in this case, B will (might) occur.

The wedding planner

- The promising but still uneven relationship between uncertainty and fuzziness in natural language and argumentation schemes needs a systematic development
- Ingredients:
  1. a classification of uncertainty/fuzziness types
  2. a characterization of the uncertainty/fuzziness types relevant to each argumentation scheme
  3. a formalism for the representation of uncertainty/fuzziness assessments (of various types) in actual arguments, i.e. in instances of argument schemes;
  4. a mechanism to derive an uncertainty/fuzziness assessment for the conclusion of an argument from the assessments concerning the premises and the applied scheme.

Classifying uncertainty types

- Searching "ontology of uncertainty“ on the web the most authoritative link found is by W3C
- The page is entitled:
  W3C Uncertainty Reasoning for the World Wide Web XG
  UncertaintyOntology
- Followed by the note:
  This is an archive of an inactive wiki and cannot be modified.
- Did the W3C surrender to this challenge?

Excerpts from W3C ontology

- **Sentence** - an expression in some logical language that evaluates to a truth-value (formula, axiom, assertion)
- **World** - the world about which the Sentence is said
- **Uncertainty** - a statement about the uncertainty associated with the sentence
- **Uncertainty Nature** - whether the uncertainty is an inherent property of the world or is a lack of information
  - » **Aleatory** - the uncertainty comes from the world; uncertainty is an inherent property of the world
  - » **Epistemic** - the uncertainty is due to the agent whose knowledge is limited, especially for a machine agent

Excerpts from W3C ontology

- **UncertaintyType** - classification of uncertainty
  - » **Ambiguity** - the referents of terms in a sentence to the world are not clearly specified and therefore it cannot be determined whether the sentence is satisfied, see also http://en.wikipedia.org/wiki/Ambiguity
  - » **Empirical** - a sentence about a world (an event) is either satisfied or not satisfied in each world, but it is not known in which worlds it is satisfied; this can be resolved by obtaining additional information (e.g., an experiment)
    - – **Randomness** - sentence is an instance of a class for which there is a statistical law governing whether instances are satisfied
  - » **Vagueness** - there is not a precise correspondence between terms in the sentence and referents in the world, see also http://en.wikipedia.org/wiki/Vagueness
  - » **Inconsistency** - there is no world that would satisfy the statement
  - » **Incompleteness** - information about the world is incomplete, some information is missing

Excerpts from W3C ontology

- **Uncertainty Derivation** - how the fact about uncertainty was derived
  - » **Objective** - derived in a formal way, repeatable derivation process
  - » **Subjective** - subjective judgement, possibly guess

Excerpts from W3C ontology

- **Uncertainty Model** - mathematical theories for the uncertainty types
  - **Probability**
  - **Fuzzy Sets**
  - **Belief Functions**
  - **Random Sets**
  - **Rough Sets**
  - **Similarity Models**
  - **Preference Models**
  - **Trust Models**
  - **Combination Of Several Models**
    - **Fuzzy Sets And Probability**

Excerpts from W3C ontology

- **Properties**
  - **hasUncertainty** - sentence S has uncertainty U
  - **saidAbout** - sentence S is said about world W
  - **saidBy** - sentence S was said by agent A
  - **nature** - uncertainty U has nature N (either aleatory or epistemic (lack of knowledge)
  - **uncertaintyType** - uncertainty U is of type T
  - **uncertaintyModel** - uncertainty U is modeled using the mathematical theory M
  - **derivationType** - uncertainty U was obtained by derivation of type D

Another uncertainty ontology

- Another top link returned by the search points to a paper:
  "Ontology of Scientific Uncertainty: Methodological Lessons from Analyzing Expressions of Uncertainty in Food Risk Assessment"

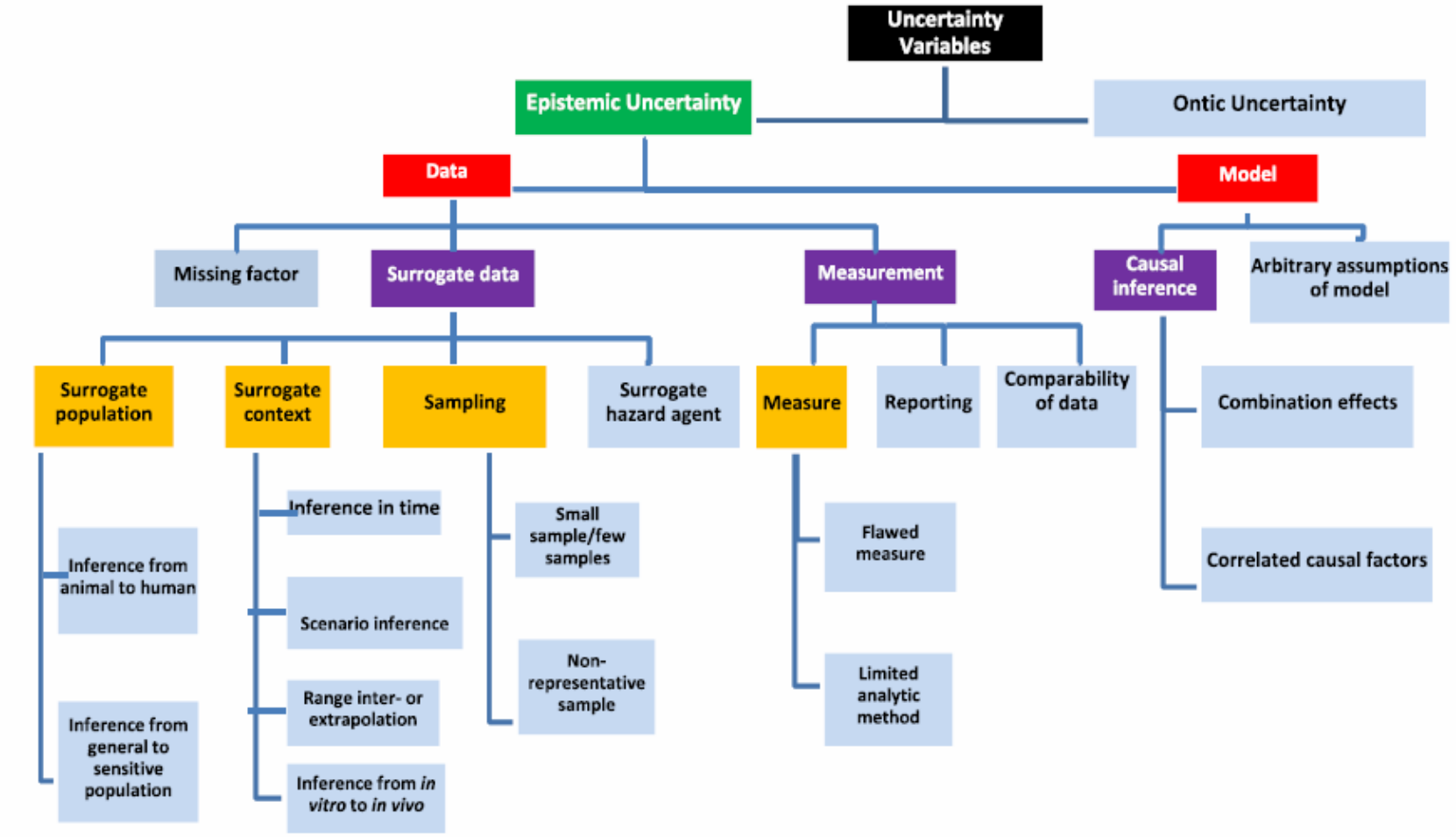
Another uncertainty ontology: origin rather than form

Table 1. Decision Tree for Uncertainty Taxonomy Coding

Is it uncertainty that is irreducible? OR *Is it that new information can resolve*	Ontic Uncertainty/Variability				
	Epistemic Uncertainty *Is it due to the absence of good data about the hazard?* OR *Is it due to the way the model is built?*	Model *Is it due to arbitrary model assumptions?* OR *Is it due to some problem in our causal understanding i.e. what generates the hazard?*	Arbitrary assumptions of model		
			Causal inference *Is uncertainty due to ignoring synergism (combination effects)?* OR *Is it due to the inability to separate the effects of related causes?*		Combination effects
					Correlated causal factors
		Data *Is it due to the complete absence of data* OR *Is it due to the lack of the exact kind of data we need and the fact that we have to use proxies (surrogates)?* OR *If we have the right kind of data, does some quality of the data create uncertainty ?*	Missing factor		
			Measurement *Data from different sources are incomparable and they point in different directions.* OR *We don't know enough about how it was measured to trust the data.* OR *Is it due to how it is measured?*	Comparability of data	
				Reporting	
				Measure *Was the measurement poorly done?* OR *Does the methodology used in measuring have inevitable limitations?*	Flawed measure
					Limited analytic method
			Surrogate Data *Is it a sampling problem?* OR *Is there a discrepancy between the hazard we have data for and the hazard of interest?* OR *Is there a discrepancy due to context?* OR *Is there a discrepancy between the population of interest and the population we have information on?*	Sampling *Was the sample too small?* OR *Was it selected improperly?*	Small sample size /few samples
					Non-representative sample
				Surrogate hazard	
				Surrogate context *Are the data from the wrong context?*	Inference in time
					Scenario inference
					Range inter- or extrapolation
					Inference from in vitro to in vivo
				Surrogate Population *Are the data from the wrong population*	Inference from animal to human
					Inference from general to sensitive population

Another uncertainty ontology: origin rather than form

Figure 2. The structure of the ontology of uncertainty based on content



Ontology of uncertainty

- A prerequisite for many important developments including argument mining
- A formidable philosophical and technical challenge
- A psychological challenge too?
- Shall we continue to build theories and systems with shaky foundations (if any)?

Yes!*

- *If shaky experiments are meant to contribute to a better understanding of how to lay down solid foundations

A shaky experiment: some uncertainty types

Just for the sake of experiment:

- source uncertainty [U1]: to evaluate the credibility of different statements one may take into account the credibility of their sources
- uncertainty about a statement [U2]: a subject expresses a partial degree of commitment to a statement s/he makes;
- uncertainty inside a statement [U3]: linguistic uncertainty generically present in natural language statements
- derived uncertainty [DU]: arising from propagation in the reasoning process

A shaky experiment: uncertainty in argument schemes

- The scheme specification should be accompanied by an explicit account of the types of uncertainty it may involve
- The use of linguistic uncertainty expressions in the scheme (like in ACE) should be avoided within the natural language description of the scheme itself
- Goals:
  - » Explicit modelling choices
  - » Uniformity
  - » Non ambiguity

A shaky experiment: uncertainty in argument schemes

- Uncertainty types may be associated with:
  - » Premises (possibly affected by uncertainty)
  - » Critical questions (pointing out possible uncertainties)
  - » The scheme itself and its applicability (crucial but not considered in this example)
  - » The conclusion (derived from all the previous ones)

A shaky experiment: APK scheme

A may contain linguistic uncertainty

Source a is mentioned

- Major Premise: {Usually, if a is in a position to know about things in the subject domain S}[U1] {containing proposition A}[U2]

CQ3: next slide

- Minor Premise: {a asserts that A (in domain S) is true (false)}[U1;U2].
- Conclusion: {A is true (false)}[DU].
- CQ1: {Is a in a position to know whether A is true (false)?}[U1]

Different doubts about the source

- CQ2: {Is a an honest (trustworthy, reliable) source?}[U1]
- CQ3: {Did a assert that A is true?}[U2;U1].

CQ3: next slide

CQ3 of APK: what does it mean?

- CQ3: Did *a* assert that *A* is true?
- First interpretation: doubt about the fact that *a* did actually make any assertion about *A*.
  U1 uncertainty about the source of the information about the assertion made by *a*
- Second interpretation: doubt about the contents of the assertion (e.g. *a* asserted that probably *A* is true)
  U2 uncertainty: *a* might be not so certain about his/her own statement

A shaky experiment: ACE sch

Source of the causal relation could be questioned - CQ2

Uncertainty inside the the original statement (might) and CQ1

Missing critical question CQ+ similar to CQ3 in APK

- then B will occur} [U1;U3].
- [U1;U3].
- Conclusion:{Therefore, in this case, B will occur} [DU].
- CQ1: {How strong is the causal generalization?}[U3]
- CQ2: {Is the evidence cited (if there is any) strong enough to warrant the casual generalization?}[U
- CQ+: {Does A actually occur?}[U
- CQ3: {Are there other causal factors that could interfere with the production of the effect in the given case?}[??]

This basic classification is definitely incomplete

Which direction to go?

The (temporary) ostrich approach

- Simply ignore uncertainty and fuzziness (UF) by now and go ahead with the UF-free study of argumentation and natural language
- Uncertainty and fuzziness will be added on top of UF-free solutions
- Justification: Would be to be difficult to add the conceptual and technical complexity of UF to an already very complex problem
- Rebuttal: UF are so fundamental that ignoring them is like not really facing the same problem

The 'improve the schemes' analytical approach

- Develop a more systematic analysis of argument schemes with UF
- Make argument schemes more expressive and more complete
- Use the improved schemes for the study of argumentation and natural language
- Justification: a patient incremental approach starting from a widely accepted basis is the only way to go
- Rebuttal: you will get lost in the plethora of different schemes, questions, UF types without achieving real generality

The (temporary) 'only ad-hoc can work' approach

- Focus on very specific problems featuring a limited variety of schemes and UF
- Develop ad-hoc solutions for them: it is a first step and already a good result
- Generalizations will come later based on these experiences
- Justification: driven by actual application needs, may provide a lot of feedback from real cases
- Rebuttal: generalizing from ad hoc is the same as generalizing from scratch

The 'new foundations' approach

- If all the approaches considered up to now appear hopeless, we should look for something new
- An original, highly generic modeling approach to relate argumentation, natural language and UF is needed: conceptual and foundational work first!
- Justification: without suitable foundations nothing serious can be built and short-term efforts are just a waste of time and energy
- Rebuttal: Best wishes for your Holy Grail search!
  All (too) long-term efforts are just a waste of time and energy!

Really no sign?

The 'no direction in early days' meta-approach

- Leave every researcher (and every research clan) free to try her/his preferred approach
- Promote occasions of exchange between the clans
- Try to avoid that each clan has its own events and each event has its own clan(s)
- Try to avoid to reject papers just because they are "too ad hoc" or "too abstract" or "too preliminary"
- Don't require immediate results (you'll get the fake ones)
- Believe in community strength and slow progress


Uncertainty


Any question or comment will help