Trust in Intelligent Machines Workshop (TIM 2018)

A humane interface for intelligent machines

Simon Wells
School of Computing
Edinburgh Napier University

Scenario

- Increases in capabilities of intelligent machines
- Long-term trend for people to delegate decision making to machines
- Moral outsourcing
- Tendency for humans to mistrust anything different

Conflicting forces

Individual Trust

- Trust is often about our relationship to the unknown
  - If the target of trust does what we want, e.g. capable, predictable, reliable - we often don't consider trust
  - It's when we don't know some aspect of the target system that trust becomes an issue
- Individual & personal - related to a person's tolerance of & comfort with risk (& the stakes)

The Unknown

Handing Conflict & the Unknown

- How do we deal with conflict in the real world?
  - We argue, negotiate, require justification, reach agreement
- How do we deal with the unknown in the real world?
  - We find out more, ask questions, seek explanations

Argumentative Dialogue

Argumentative Dialogue & Trust

- We trust things we understand
- We understand by exploring and **explaining**
- We build confidence by **justifying**

A dialogical interaction system can support both explanatory & justificatory modes of communication between people & machines

Explaining & Justifying

- We can model dialogues as protocols and manage interactions between speakers.
- Previous work:

MAgtALO - MultiAgent Argument Logic & Opinion

DGDL - Dialogue Game Description Language

ADAMANT - A DiAlogue MANagement Tool

Argumentative Dialogue as an Interface to Intelligent Systems

1. Recognise patterns of reasoning (schemes)
2. Use schemes to instantiate arguments
3. Interact with intelligent systems via structured dialogue (explanatory & justificatory dialogues)

Problematic Approach

- Huge research challenges:
  - Natural language generation
  - Data to knowledge
  - People

Benefits

- A system that supports explanations
- A system that can justify decisions
- A system that is independent of the intelligent system
- Can be used to build trust:
  - I ask for a decision, then interrogate that decision and come to understand it. I get rid of the unknowns
- Other contexts: Legal & regulatory interaction