Lying in Real World Argumentative Dialogue Systems

Simon Wells and Al Baker

July 9, 2018

Abstract

With the increasing prevalence of artificially intelligent machines in everyday life, a trend that threatens not only to continue but to accelerate, the need to examine how people interact with these machines intensifies. Whilst the basis for much of the increased interest and utility of AI has been rooted in machine learning and neural network based systems, there are also areas of particular concern for argumentation theorists. For example, regardless of how an AI decision is made internally, should that decision be called into question, then the system should be able to explain itself, and perhaps even defend itself, furthermore, the system should be able to work with people to improve decisions, should they be found wanting.

This is in line with recent trends stemming from various regulatory and professional bodies, which have independently proposed that artificial intelligence systems be capable of explaining their decisions. This trend is found both at the supranational regulatory level, in recommendations from the European Commission, as well as at the industrial professional level, in British standards for intelligent and autonomous robots.

It would appear that many years of research into formal argumentative dialogue systems may soon result in real-world payoffs. However, thorny questions remain in relation to how our ideal, normative systems of argument and dialogue will fair when exposed to real-world motivations.

Whilst it is often assumed that the truth should, or will, always be told, this can be easier said than done, and even when achievable, can be counterproductive. In this paper we attempt to shed light on some gray areas concerning truth telling, or lack thereof, in relation to human dialogical interaction with AI systems. From this investigation, we make recommendations for the design of future, real world, applied dialectical argumentation systems.