Expositions on Logical Consistency for Machine Learning

D.G.

June 16, 2018

1 Introduction

It appears logical consistency is a key component in verifying results and can also be used as the basis space for logical reasoning as a whole. Providing a framework for such ideas such that is it is compatible with and interpretable in the context of modern machine learning techniques could yield useful results.
In this exposition I will try to delineate my ideas on consistency. I will also try to somewhat formalize it through mathematical language and finally apply the ideas on common benchmark problems (MNIST) and on more pertinent problems (Question and answering tasks) to demonstrate the broad applicability of the idea.

2 Framing the problem

There are multiple useful concepts and observations needed to understand this paper. For us to inquire into logical consistency it would be a good idea to observe certain cognitive happenings.

2.1 State Imputation

Consider the following scenario: Bob walks into his shed to discover the undesirable property of there being a wet spot on the ground. He investigate further and discover that there appears to be a small whole directly above the mark. This is in itself not a justifiable cause for the problem. If it were mid-July and weather dry, one would have to inquire elsewhere, but it happened to be mid-November and it just rained heavily yesterday. Finally, he concludes that the hole was the cause of the wet mark on the ground. How did Bill arrive at this conclusion? It appears that one can frame these thought processes in more clear sequences. For this example, one can verbalize such transitions as:

1. There was a water mark on the ground;
2. There was a hole in the roof and such holes have the capability of allowing liquids to pass through it;
3. It rained water (a liquid) recently which would cause things to be wet;
4. $\implies$ The hole let liquid from the rain pass thought it, subsequently wetting the ground.

We can re-organize this into state causal transitions. For example, observations that "Water causes things to get wet" and "Holes can cause water to pass through otherwise impermeable materials".
As you can see, there is already a high level of abstraction in such resolutions to the extent that automation isn't readily visible. For instance, how to deduce holes let liquid pass though it, that water is a liquid and most importantly, how to stray away from hand coded symbolic systems? All of these will be tackled later on but for now let us on the next underpinning idea- tautologies and entailments.

2.2 Tautologies and Entailments

The next important concept is tautologies and entailments. I refer here to the more pertinent idea borrowed from philosophy not regular English. In logic, a tautology (from the Greek word) is a formula or assertion that is true in every possible interpretation. A simple example is "(x equals y) or (x does not equal y)" (or as a less abstract example, "The ball is green or the ball is not green").

This almost capture my idea- but not exactly. The 'word' entailment describes the idea better but is too vague. In this context, I refer to tautology as: Starting from $\mathcal{A}$, all possible immediate states, *beta*, that are instrinsic in $\mathcal{A}$ are defined as tautological successors or $\mathcal{A}$ The word *intrinsic* is important here. It implies that no logical leaps or external assumptions are being made, but instead only re-phrasals. "Circles are circular" is an example because circular is defined on the idea of a circle. The concept of 'circular' is the tautological successor of circle. A false example would be "My bed-sheets are blue" because you'd be attributing information to "my bed-sheets" that is is not inherent in the definitions of "my bed-sheets". One's bed-sheets are based on definition of them having your ownership- whether they're blue or not is unwarranted attribution. An example, drawing from simple mathematical ideas, could be the fact that $(x+1)(x-1)$ is the same as x^2-1. Both objects entail the same properties but are just different representations.

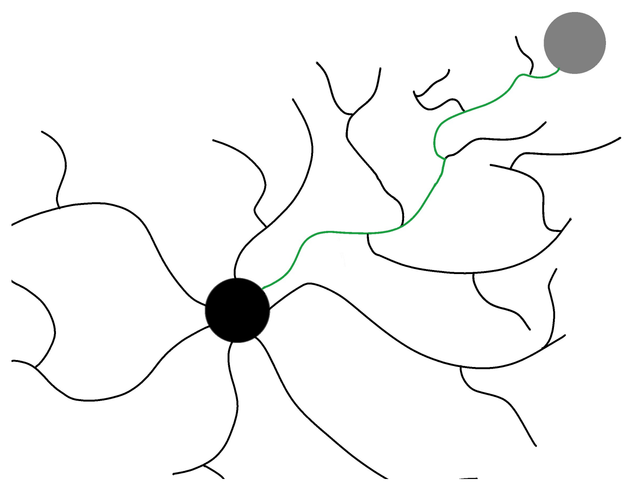

Figure 1: Abstract example.

2.3 Synthesizing

Now we have these two concepts distinguished, a framework based on a synthesis of the two is ripe for the making. Figure 1 demonstrates the structure on an abstract level. Here teh black circle denotes what we refer to as $\mathcal{A}$ in the previous section and the grey circle as β. To re-iterate; $\mathcal{A}$ is our starting idea and our goal it to find a route throught space (using tautologies) to reach β. We denote the green line connecting the two as the solution.

3 Proposed Solutions

4 Experiments