From: **Peter Latham** pel@gatsby.ucl.ac.uk
Subject: Re: Curiosity and the dilemma (eLife)
Date: December 21, 2020 at 8:38 AM
To: Erik Peterson erik.exists@gmail.com
Cc: timothy Verstynen timothyv@gmail.com

PL

No problem. Good luck with the article! I'm looking forward to a version that I can read. ;)

P

> There are a few ideas to setup, and which to do when has been a struggle. I'm not sure the its the RL problem Maybe it is.
>
> The rest of your advice - spot on.
>
> Your final counter example - hmmm. I don't think I agree, but need to play a bit. I had been assuming evolution would "save" us from such bad rules and so its an unexplored space. There's no saving us from mathematicians. :)
>
> But, it is time to leave off. Thank you for the help and time. It really was not at all required of you, and I do immensely appreciate it.
>
> (e)
>
> > On Dec 21, 2020, at 12:39 AM, Peter Latham <pel@gatsby.ucl.ac.uk> wrote:
> >
> > I admit to not having read the abstract and intro until recently. But I just reread it, and it didn't help much. Basically, it said that you have another solution to the explore/exploit tradeoff. I ignored the fact that you called it curiosity, since people are always giving names to what turns out to be standard algorithms.
> >
> > Here's what I would have said in the intro:
> >
> > - agents almost always face an explore/exploit tradeoff.
> > - the classic solutions to this problem have been X.
> > - ours is different because of Y.
> >
> > And in results, I would have started by setting up the RL problem -- not everybody (including me) is an RL expert. Then I would have described two things: how the agent decides whether to ignore reward and explore, and what the explore strategy is. Possibly in the opposite order. But with a minimum of abstract definitions, and with lots of intuition. And good notation.
> >
> > By the way, presumably your scheme works for my function, which I'll
> > modify to
> >
> > f(x, theta) = theta/(2 + sigmoid(x)/10^100)
> >
> > If that's true, then it seems like something is wrong, because it's too simple.
> >
> > P
> >
> > > You read the abstract and intro? Or you began, as I do, in truth, with the results?
> > >
> > > Or perhaps even with the Intro/Abstract read you're - putting all courtesy aside - saying wtf, why is this all happening?
> > >
> > > (e)
> > >
> > > > On Dec 20, 2020, at 4:22 PM, Peter Latham <pel@gatsby.ucl.ac.uk> wrote:
> > > >
> > > > Basically, the first couple of pages of Results. You're talking about curiosity, information and memory, but I have only a vague idea what's going to be done with them -- and the vague idea comes from common sense, not the paper. I can guess that they'll be used later, but to get to that I have to first memorize a bunch of definitions. You need to set up the big picture first, so that I'm motivated to pay attention.
> > > >
> > > > P
> > > >
> > > > > We've come this far. Would you point out the first three times?
> > > > >
> > > > > (e)

> On Dec 20, 2020, at 3:37 PM, Peter Latham <pel@gatsby.ucl.ac.uk> wrote:
>
> Hi, Erik.
>
> OK, that was somewhat satisfying. Although I admit that I haven't absorbed any of the technical details, at least I know what's going on. I glanced at the paper, and although it's probably possible to extract all of that, it's hard -- and impossible in the first few pages of results. Basically, the paper consistently violates my number one rule of writing, which is: the reader should never, ever, ever have to ask "why am I being told this?". If I have to ask myself that more than about three times in a row, I stop reading. I'm guessing other people are similar.
>
> Best,
> Peter
>
> > That is correct sir!
> >
> > > It doesn't matter which one you use; your scheme will work with all of them. (Or maybe most of them.)
> >
> > All of them work, I think.
> >
> > No idea how one could ever prove that. I don’t have a counterexample, at least.
> >
> > There are weird corner cases: MCTS and Rmax, who have their own exploration policy built in. One can’t remove it without destroying the alg itself. Our explore policy would ‘feed' these just fine, but it seems kind of, I don’t know, awkward. Should work though. Need to test it.
> >
> > (e)
> >
> > > On Dec 20, 2020, at 3:06 PM, Peter Latham <pel@gatsby.ucl.ac.uk> wrote:
> > >
> > > OK, so here's what I have:
> > >
> > > - No matter what action you take, you run some RL algorithm that updates parameters.
> > >
> > > - Sometimes you try to maximize reward; other times you explore.
> > >
> > > - You have a rule that tells you which one to use.
> > >
> > > - If you explore, you have a rule for how to do that.
> > >
> > > - There's also an RL algorithm, but presumably it doesn't matter which one you use; your scheme will work with all of them. (Or maybe most of them.)
> > >
> > > Is that correct?
> > >
> > > P