I think that "learning language like a baby" is a fantastic and important research area ... I'm just not (at this moment) seeing how to boil it down into a crisply-defined *challenge* that neatly gauges incremental progress ....

"Doing X" is straightforward to measure from a challenge-problem context, whereas "Learning X" is harder to measure in a challenge-problem context, because eager competitive contestants can always program most of X into their system and then make their system kinda-learn the rest.... Diamandis's X-Prize Foundation has asked me for help with formulating an AGI X Prize multiple times over the years, and I never had anything great to suggest for precisely this reason.   Of course a prize for "achieving human level AGI" would make
sense-- but that's such a big achievement that if you get there, the
prize will be the least of your worries anyway!   It's measuring
incremental progress in a rigorous and cheating-proof way that's so tricky...

About how babies learn language. Clearly it's a lot about embodiment and social interaction, right? You may have read Tomassello,
"Constructing a Language"?   (not AI, cognitive science, but
good...) ...   And I think rich perceptual stimuli and some degree of
motoric affordances are important.   So to really emulate or
understand learning language like a baby, I suspect it's necessary to go robotic (though, certainly, not necessarily HUMANOID-robotics).
Joscha may disagree on this point, I'm unsure....   The point is you
need a rich stream of perceptual data, whose interrelationships can ground the interrelationships between linguistic constructs; and you need actions to be taken based on this perceptual data, to give grounding for the structure of sentences (which is action-based at the
base, with the VERB at the center of the sentence, etc.).   In theory
this could all be done in a virtual world (I mean: in theory we might all live in a virtual world!!), but it might need to be a lot more data-rich than Minecraft...

The Europeans are pushing in this direction with iCub, but very slowly, as always w/ these massive multi-university multi-nation
government boondoggle projects. Aldebaran Robotics was doing
something in this direction, but that research group was shut down
when they were acquired by Softbank a year or two ago. Google,
oddly, seems not to be doing this sort of thing (yet) --- even though they have some great folks doing computational linguistics (including unsupervised learning of syntax from corpora) and they have just bought a raft of robotics companies...

So my own feeling is that to make progress on "learning language like a baby" you want to use a simple robot that needs to do stuff in an environment, and needs to learn language to achieve its goals in that
environment....   Could be a simple rolling robot with a camera,
microphone, speaker and arm, moving around in a robot-lab environment (but NOT in a "playroom" denuded of diversity of objects and events)... or could be a simple humanoid...

One idea would be to go back to the idea of a child IQ test. The

challenge would be to make a robot that could pass some preschool IQ

tests.   Granted, this would not focus efforts entirely on learning,

because people could hack stuff just oriented toward the specific tests. But by making the tests more and more unpredictable, one could

make this sort of hacking harder....   I think it needs to be a robot

for this, because if you abstract the preschool IQ tests into a simplified digital form, they become too easy (they often have to do with the intersection of vision, movement and cognition).... In the actual physical-world form, the AI has to understand the relation of the test to the physical environment, etc. ...

Of course, one could make both a Minecraft and a robot version of the same preschool IQ tests, and empirically see how well ability to pass the preschool Minecraft IQ tests, helps in terms of conferring ability to pass the robot preschool IQ tests...

One thing that would interest me -- and Joscha -- would be working toward embodied agents that could pass these IQ tests **via learning

language in an embodied way** ....   This might not be the shortest

path to making agents that could pass the preschool IQ tests, but it would be the most interesting way with the most long-term promise...

-- Ben

On Mon, Sep 7, 2015 at 8:43 AM, jeffrey E. <jeevacation@gmail.com> wrote:
> hilbert did questions ? im open for ideas   looking to encourage work
> based on the concept of coherence.   or sense making modules. . babys
> learn language. many of them . figure out how the baby does it.
>
> On Sun, Sep 6, 2015 at 8:16 PM, Ben Goertzel ███████████████ > wrote:
>>
>> Hi,
>>
>> I agree w/ Joscha's caution about discrimination tasks: They can be
>> often be solved rather well, but in devious ways, by statistical
>> supervised learning algorithms. Suppose you pose a linguistic
>> discrimination task of some sort -- and a supervised learning
>> algorithm, trained on a mass of data, can solve it with 97% accuracy.
>> The algorithm's pattern of errors may indicate to YOU, intuitively,
>> that it doesn't really understand what's going on. But then, it may
>> be that the average person solves the task with only 95% accuracy,
>> though with a different pattern of errors that indicates intuitively
>> they have a different kind of understanding...
>>
>> I like the idea of a language learning challenge, but posing it
>> properly seems tricky.   As soon as something becomes a "challenge",
>> one has to worry about protecting against various subterfuges
>> (deception, once again!).   Suppose one poses a challenge to learn a
>> language from an un-annotated corpus of texts. OK, but then some
>> nefarious clever person can try to solve this using an algorithm
>> whose parameters were all carefully tuned via analysis of an annotated

>> corpus in that same language. And these parameters may be quite
>> complex structures. The winning approach would then not be able to
>> work on another language for which there was no large annotated
>> corpus (no Penn Treebank analogue, etc.).
>> It seems that challenges are easier to formulate for engineering
>> breakthroughs than science breakthroughs...
>>
>> Here is one idea, off the top of my head.... Perhaps at least it can
>> stimulate thoughts .... This is not about language learning, though,
>> it's about recognizing and generating coherent, meaningful language..
>>
>> 1)
>> Show human subjects some videos of game characters carrying out
>> certain sequences of behaviors in a video-game environment
>>
>> 2)
>> For each behavior-sequence B, ask the human subjects to generate some
>> textual instructions, that would enable the reader to emulate
>> behavior-sequence B (even if the reader had not seen the videos)
>>
>> 3a)
>> Ask the AI to figure out which textual instructions would actually
>> work, for each behavior-sequence B
>>
>> 3b)
>> Ask the AI to actually generate textual instructions, based on
>> behavior-sequences (then the judgment is whether people, when
>> following, the AI's instructions, actually carry out the appropriate
>> sort of behavior sequence)
>>
>> Note that 3a and 3b both measure "coherence" in a concrete and
>> obviously meaningful way...
>>
>> ...
>>
>> I remember seeing some NL generation challenge vaguely like this a
>> few years ago, but don't have the link handy. Ruiting will probably
>> be able to find the reference if it's of interest...
>>
>> ...
>>
>> For language learning, the only good way I can think of to make a
>> challenge would be to use languages for which there are no annotated
>> corpora. So, the challenge would be to take some unannotated text
>> (or speech) from an arbitrary human language (could be an Australian
>> aboriginal language, or an African language, etc.), and then figure
>> out how to generate grammatical and coherent utterances in that
>> language. This is pretty hard obviously. If someone chose to
>> "cheat" by building annotated corpora or rule-bases for every obscure
>> language in the world, at least they would be doing the world a big
>> service along the way ;-D
>>

3

>> Interesting thought-direction, anyhow... !
>>
>> -- Ben
>>
>>
>>
>>
>>
>>
>>
>> On Mon, Sep 7, 2015 at 4:23 AM, jeffrey E. <jeevacation@gmail.com> wrote:
>> > I dont want statistical modeling   you and ben for years have stated you
>> > wanted to put an avatar , and hope it can do things a 2 year old can do.
>> > the challenge is learning a language.   different that moving blocks in
>> > a
>> > video game.
>> >
>> > On Sun, Sep 6, 2015 at 2:38 PM, Joscha Bach ███████████████████ >
>> > wrote:
>> >>
>> >> This challenge idea is excellent; I really love it!
>> >>
>> >> > first draft. of the Chomsky Challenge.   . Produce a non- living
>> >> > system that can be put into an environment for a while   and ---
>> >> > 1 . be
>> >> > able to discriminate   language from noise. .   prize . a 1 dollar
>> >> > bill
>> >> > signed by Noam and 100k.
>> >>
>> >> What is the system allowed to have when it starts? We would need
>> >> to define the environment, for instance text based or audio, or
>> >> movies/youtube.
>> >> Once
>> >> the contestants know the environment, they can use standard
>> >> machine learning methods to discern entropy in the signal, and
>> >> separate language-like noise from non language-like noise. Google
>> >> does this pretty well, and automatically (but not perfectly)
>> >> sub-title videos in a number of languages.
>> >> I imagine you want to go beyond that?
>> >>
>> >> > . 2. be able to discriminate coherent sentences   from non (
>> >> > we
>> >> > provide 10 test sentences).
>> >>
>> >> I suspect that this is harder, Noam might point out that a lot of
>> >> grammatically well-formed sentences used in politics are not
>> >> coherent
>> >> ;-)
>> >>
>> >> > prize a 10 dollar signed Chomsky bill , and 500k. 3. a
>> >> > language learning module.
>> >>

4

>> >> Build a system that is able to learn a new language without
>> >> hand-coding, and translate sentences from this language into
>> >> English and back?
>> >> Excellent!
>> >>
>> >> > 20 dollar bill signed and 1 million, 4. a sense making module
>> >> > that can
>> >> > understand meaning inference. . etc. the non recommendation
>> >> > recommendation.   . ie the student has a nice family. etc. a 100
>> >> > dollar
>> >> > signed bill and 10 million dollars. ? ---
>> >> >
>> >> > lets also do a minsky challenge and if you want martin a NOVAK
>> >> > challenge.
>> >>
>> >> Yes! Let us ask Marvin and Martin about the biggest unsolved
>> >> problems in their field.
>> >>
>> >
>> >
>> >
>> > --
>> >   please note
>> > The information contained in this communication is
>> > confidential, may be attorney-client privileged, may
>> > constitute inside information, and is intended only for
>> > the use of the addressee. It is the property of
>> > JEE
>> > Unauthorized use, disclosure or copying of this
>> > communication or any part thereof is strictly prohibited
>> > and may be unlawful. If you have received this
>> > communication in error, please notify us immediately by
>> > return e-mail or by e-mail to jeevacation@gmail.com, and
>> > destroy this communication and all copies thereof,
>> > including all attachments. copyright -all rights reserved
>>
>>
>>
>> --
>> Ben Goertzel, PhD
>> http://goertzel.org
>>
>> "The reasonable man adapts himself to the world: the unreasonable one
>> persists in trying to adapt the world to himself. Therefore all
>> progress depends on the unreasonable man." -- George Bernard Shaw
>
>
>
>
> --
>   please note
> The information contained in this communication is

> confidential, may be attorney-client privileged, may
> constitute inside information, and is intended only for
> the use of the addressee. It is the property of
> JEE
> Unauthorized use, disclosure or copying of this
> communication or any part thereof is strictly prohibited
> and may be unlawful. If you have received this
> communication in error, please notify us immediately by
> return e-mail or by e-mail to jeevacation@gmail.com, and
> destroy this communication and all copies thereof,
> including all attachments. copyright -all rights reserved

--

Ben Goertzel, PhD
http://goertzel.org

"The reasonable man adapts himself to the world: the unreasonable one
persists in trying to adapt the world to himself. Therefore all
progress depends on the unreasonable man." -- George Bernard Shaw

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN" "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="1.0">
<dict>
        <key>date-last-viewed</key>
        <integer>0</integer>
        <key>date-received</key>
        <integer>1441588557</integer>
        <key>flags</key>
        <integer>8590195717</integer>
        <key>gmail-label-ids</key>
        <array>
                <integer>7</integer>
                <integer>27</integer>
        </array>
        <key>remote-id</key>
        <string>540254</string>
</dict>
</plist>
```