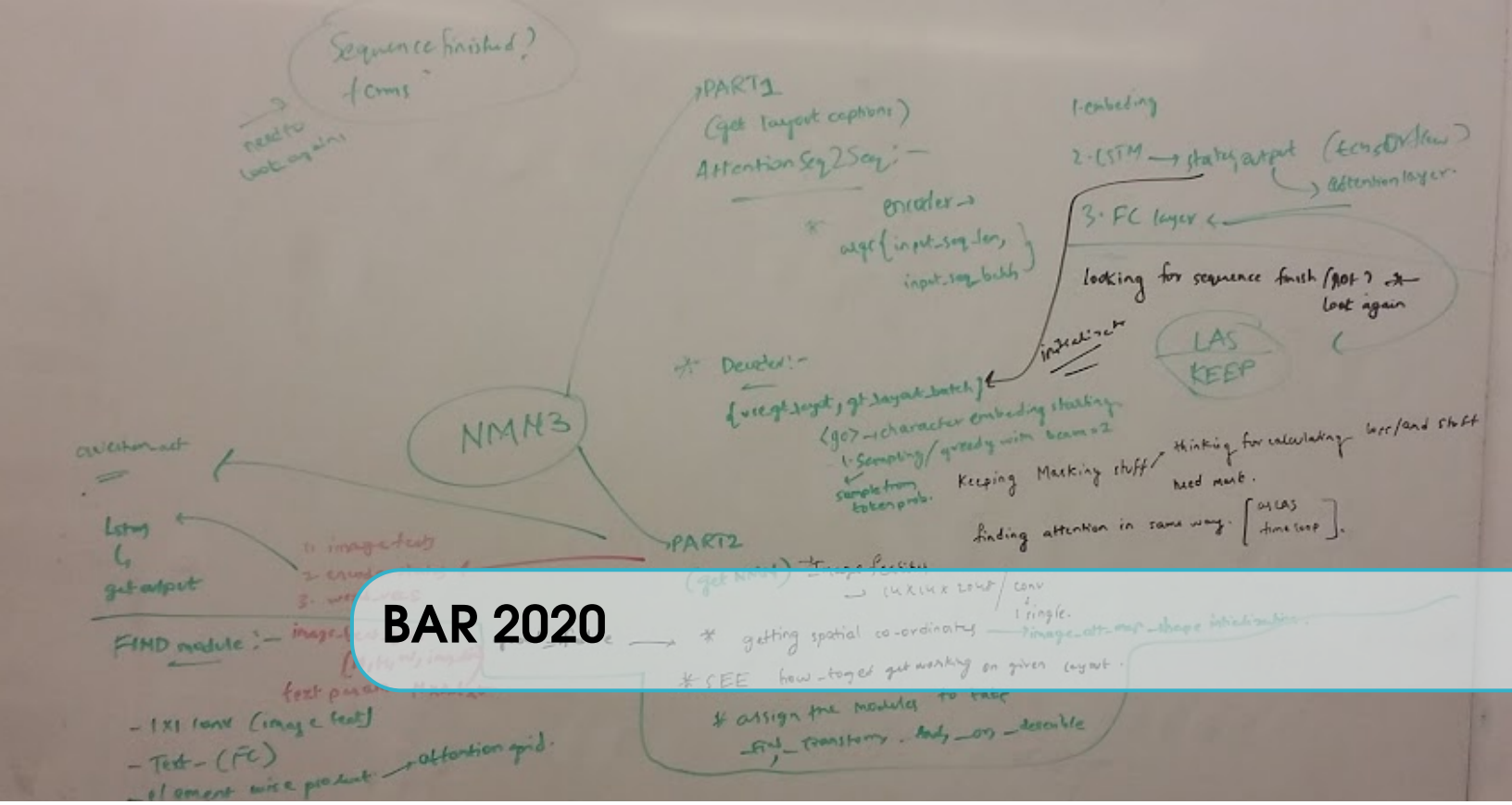
BlackBox Annual Report 2020

Image Source: University's Instagram page

This year's report was edited by Sai Krishna Rallabandi with the following contributions:
Spoken Language Understanding - Akshat Gupta
Generative Modeling - Khyathi Raghavi Chandu
Media Wiki - Xinjian Li
Notion - Peter Wu



A typical question we get from prospective students is to describe our research lab[1]. In my view, our research group can be seen as a holding company with multiple research directions each being pioneered by one or more members. Our structure is a mix of two approaches: (1) approach similar to other holding companies such as Berkshire Hathaway where each division is headed by their CEO, and (2) modern approach similar to ARK Invest where the individual divisions seamlessly collaborate with each other. The vision of our research lab is to bring language technologies closer to humans. Currently we are interested in 5 cornerstone research areas which we deem as necessary conditions to realize our vision - (1) Code Switching (2) Language Technologies for low resource and unwritten languages (3) Computational Ethics (4) Generative Modeling and (5) Explainability. We have been working for years on some of these areas while we are in the early stages in others. Having said that, our unique perspective - one based on first principles - is reflected in our research across all of these areas.

Motivation

There are two main motivations for beginning the process of annual reviews for our research lab: (1) A report similar to what we have started with this

[1] Image source: Khyathi's scribbles from board in our office

version is something me and some of the peers longed to read during their undergrad lives. While there are a number of blog posts by PhD students which provide a snapshot, a report from the research group as a whole provides the evolution of thought process of the cohort. (2) *Knowledge is hyper local*. Some of the best research practices and insights were usually subject of in-person corridor discussion among research students which facilitated exchange of knowledge as well as tactics. We believe a beneficial step would be to periodically document them and a report like this provides a neat platform for the same.

Road map

I begin this report by penning down my thoughts on the importance of higher level thinking for graduate students. By higher level thinking, I imply the ability to look beyond the usual zeroth order levers - like time management and number of working hours - that students typically employ as their KPIs. I anticipate that this part of BAR will be written by a different member each year going forward. We then present a brief overview of our current research in two of the five core research areas of our interest - (1) Language Technologies for low resource and unwritten languages and (2) Code Mixing. We then review two tools which we found productive to enhance research workflows - Media wiki for logging and Neo4j for visualization. We close this report by tabulation of our compute resources.

Highlights

- Amrith Setlur's paper[SPB20] was nominated for Best Student paper at Interspeech 2020.
- Our voice building tool was employed to provide information about COVID19 to people who may not be proficient in reading. This includes models in 650+ languages (https://www.dictate.app/synthesize). Contact Xinjian Li for details.
- Shruti Palaskar and Shrimai Prabhumoye led the SCS, CMU effort for PhD Applicant Mentoring. Details here: https://www.cs.cmu.edu/gasp

Importance of Higher Order Thinking

It has increasingly become clear to me that the research students across the planet are similar in more ways than they are diverse: All of us have similar academic backgrounds, similar insecurities, similar dreams about the kind of research we wish to carry out. What's diverse is the research experience, hence the popular rhetoric about research - you know, the one that implies a stressful period in research life. Admittedly, a lot has got to do with epigenetics - the advisor, peer group, the research environment as a whole, etc. But since we pride ourselves in looking at things from first principles, let's take a look at this aspect in more detail. And while we're at it, lets also discuss what we can do about it. My peers and I have been looking at ways to optimize our research experience and the common denominator upon loooong discussions appears to be this thing I refer to as higher order thinking. If you are reading this, you are interested in thriving and hence you perhaps already are applying this, but let us dissect it.

To understand this, consider a scenario: Imagine a student(lets call them Sharma) joining the research group. Since I am at LTI and the most typical entry is into a Masters program, I will consider this student as a Masters student joining during Fall semester. Sharma takes ambitious courses, the semester becomes very heavy very soon leading to all-nighters and coffee, followed by minimization of social life. After this, it appears Sharma goes through one of two trajectories:

- (1) $Sharma_{survivor}$ starts looking at (and blaming) their time management skills, stops attending any additional stuff such as seminars, events and is alwaaaays working. Does this variant of Sharma sound familiar? Now there is absolutely nothing wrong with working long hours if Sharma is enjoying the process. All of us work long hours. However, if they aren't, this quickly turns into a vicious cycle of sleep deprivation and other things. The cycle repeats itself the following semesters and into research as well. What's worse, $Sharma_{survivor}$ blames themselves for this. I believe this is where the rhetoric originates from.

- (2) $Sharma_{thriver}$ makes very subtle changes and continues on a blissful

path regardless of the work commitments. There sure are periods of stress, but $Sharma_{thrivers}$ seem to be able to spring out of these periods consistently.

We are interested in understanding what led $Sharma_{thriver}$ to thrive where $Sharma_{survivor}$ found it frustrating. I tend to believe that most of us are in the spectrum and therefore it helps to study the mechanics of these classes more deeply than just labeling them. There is a huge body of ongoing work investigating the role of personality traits such as resilience in this[AWM11; BAW11] and broader correspondence to mental health [HZW15]. The fascinating aspect of these studies is that interventions seem to help [BPB10; PMD98; Rob+15] and consistently show improvements across domains from physicians[Soo+11] through patients[Min+13] and also in sports[Guc+11].

One of the works that really got me intrigued is an investigation targeted at well-being[Meh+19]. The researchers hypothesized a triad of hope, world-view, and self-esteem as core contributors to mental health. Such an approach is interesting from two perspectives: (1) The variables are at a sufficiently higher level so that conclusions can be generalized to other domains (2) The variables are more or less epigenetically controllable. It appears there is a similar higher level triad in the context of academic research - one characterized by the control variables Health, Career and Finance. Remember the subtle changes I mentioned $Sharma_{thriver}$ makes? The changes seem to be almost always related to these variables. In other words, $Sharma_{thriver}$ appears to have mastered the art of pulling these levers to maneuver through a tough scenario. Note that I employ the terms in a broad sense, which is an oversimplification. For instance, Health encompasses physical health as well as mental health both of which are in turn contextually dependent on several other variables.

It seems tempting to ignore the control variables Health and Finance for the duration of one's academic life. In principle, one might interpret this as single minded focus toward career. However, in practice this seems shortsighted. To comprehend why, let us take part in a thought experiment. I will personalize this thought experiment to Pittsburgh. I also would like

to point out that this is just a scenario I am painting to make the case for optimization of three variables as opposed to maximization of one. I urge you to take the deductions with a grain of salt. Consider Sharma has chosen to maximize the career variable by completely ignoring the other two variables. To maximize their potential, Sharma spends all day working. Once the semester completes, Sharma chooses to stay back and work during the break as well. This pattern repeats for the next 3 semesters. Let us visualize what's happening with Sharma biochemically.

I will choose one example component to take us through this visualization - Vitamin D. I like demonstrations involving Vitamin D since it is a commonly overlooked vitamin. We make Vitamin D mainly from sunlight. Although it can also be found in foods such as milk, the quantity is orders of magnitude less than our requirement. Our skin interacts with UV-B radiation from sun and synthesizes Vitamin D. However, UV-B does not even enter our atmosphere in areas beyond 37 degrees North latitude during winters and therefore insolation[2] during winters will not promote its synthesis[WKH88]. While Vitamin D is generally known to regulate our mood, it also has hormonal functionality[Mor05], controls hundreds of genes in our body[NSC19] and performs crucial functions such as DNA repair [Gra+16]. Deficiency of this super vitamin has been linked to co-morbidity[Dem+19], respiratory disorders[HRM76] and associated with more than 160 pathways linked to cancer, autoimmune disorders and cardiovascular disease [HSH13].

Wait, what? Does this mean anyone who lives above 37N latitude is prone to these things? Well, not really. I apologize if I alarmed you, but we can bank on evolution to handle this issue. At this point, I would like to remind you of hibernating animals, specifically bears which remain inactive for months during winters. Thankfully, evolution is smart enough to handle this scenario for us. Vitamin D is fat soluble and biochemistry is fully capable of metabolizing it [Wam+13] from stored fat. This phenomenon has been demonstrated across species, for example in polar bears[Ves+11][3] and has been corroborated by studies from weightloss[Mas+11] in humans. Additionally, spending as little as 15 minutes per week in sun light has been

[2] INcoming SOlar radiaTION - the amount of solar radiation reaching a given area.

[3] Polar bears significantly increase their calorific intake to boost their fat stores anticipating this[Ves+11]

shown to boost its levels. So, chill.

But, lets return to Sharma. Given the sedentary lifestyle of today combined with unforeseen geo-dependent health issues like the aforementioned scenario, there is a good possibility that Sharma does develop Vitamin D deficiency. In fact, numerous studies have documented this behavior in immigrant students[Awu+98; LJ18; Mee+11]. Even if Sharma's immune system - which by the way is also regulated by vitamin D[Ara11; Bae+10; BMR11; Hew12] - is hyperactive and appears to takes care of day to day health, it is not difficult to see that the serotonin level are affected[Par98], leading to a reduction in perceived 'productivity'. Moreover, while biochemistry can compensate for short term lapses, it seems prudent not to subject ourselves to chronic deficiency, leading to a proverbial 'winning the battle but losing the war' of life.

A much better strategy seems to be manipulating the three factors to optimize the academic research life. An obvious example which all of us already are aware of is one where we expend our Health control variable(in the form of sleep) during the final stages of the semester. Another example of this would be to expend the Finance factor to outsource the jobs of grocery shopping, cooking and cleaning especially right before and in the beginning stages of the semester so that you can either (1) design your project team composition[4] extra strategically and/or (2) take that extra brain storming session converting the class project from good to kickass and/or (3) implement that one extra approach that gives the project a publication potential and/or (4) attend those talks by people from other departments firing away your brain in different directions and/or (5) connect with that undergrad student with whom you can form a symbiosis where you get to explore a new research direction and the undergrad gets exposed to research and/or (6) attend the events by any/all of the wonderful organizations within CMU like Eberly, Career and Professional Development Center, Global Communication Center and/or something else I haven't mentioned here.

An important thing to realize is that the three control variables are not orthonormal. In other words, each affects the other two. For instance, working out and meditation, though seem to correspond to only Health, are hormetic

[4] the best teams I found appear to comprise one or two PhD students, one or two research Masters and one or two professional Masters students

stressors[Ji+10] to our body and have been shown to turn on the executive parts of our brain, improving focus. Meditation has been shown to affect multiple pathways directly improving functional connectivity in the brain regions[Tar+17] as well as shown to improve telomere[5] lengths. A typical argument I came across from high performance students against meditation is about its mellowing affect: the consensus seems to be that meditation calms someone down and therefore they might lose their edge. One of the best counter arguments I heard and agree with comes from Tim Ferriss in a podcast with Dr. Peter Attia[6] where he describes the practice of meditation as the wrapping on a side of a double edged knife that protects us from getting hurt, while allowing us to employ the knife to accomplish useful tasks.

Personally, I found it trivial to balance one of the control variables at any given point, relatively manageable to balance two of them consistently but extremely difficult to balance all of them consistently throughout the year. I mentally have this metaphor of a triple ball stack balancing act - the key is to be okay with having one of the balls in the air at any given point. Trying to catch all three of them might result in all of them ending up on the ground. The positive thing - I don't think we need to balance all the three on a daily basis. As long as we have a sense of harmony between the control variables on a per year basis, we are good.

Strategy vs Tactics

While the aforementioned strategic thinking helps plan schedule before a semester, inevitably things happen during the semester and we are faced with difficult scenarios. Typically such a scenario spans just one week during a 14 week semester and has the potential to derail the entire semester if not handled well. Therefore, it is also useful to have a set of tactics ready. The most common scenario I have encountered with myself and peers is that of being overwhelmed. The state of overwhelm is experienced when things seem just out of reach or beyond reach. Our brain is wired to have a feeling of control and serenity. So when we lose control, the reptilian part of our brain kicks in and starts stress. This was natural way for us to cope

[5] Telomere length is an important biomarker related to metabolic aging and senescence [Epe+09]
[6] https://peterattiamd.com/timferriss

with uncertain scenarios for instance thousands of years ago. Default reaction in such scenarios is to try to get out of the situation as soon as possible, since our brain wants to experience control and serenity. But there is merit in trying to control the situation and adapt. A practice I found very helpful to deploy in the scenarios is to realize that my primary source of fuel is not the number of hours put on a task, but the amount of energy. I found that the trick of changing the state of mind helps replenish the energy levels and power through the scenario. Change of state can be manifested in a lot of ways: Talking to someone, working out, naps, cold showers, meditation, doing things very slowly(for instance, typing with one hand).

All we need is a thread

I would like to end this part with an observation about the cover picture which personally influenced me quite significantly. The cover picture is a snapshot of the first GPU machine we built in our research lab called K20. After assembling the entire machine we noticed that we got the wrong type of cooling fan for CPU. Slightly disappointed, we were wrapping things up when suddenly the engineer in us wanted to see what happens if we attempt a small hack. The blue thingy you can see on the cover page is not a wire, but just a thread which we used to hold the components in place. This machine is stable till date and had the least number of issues. I believe there is a neat takeaway from this incident. When things seem astray, all we need is a thread to hold on to. I tend to believe that this thread is the narrative we have with ourselves. What's more - it need not even be real. Placebo works. As long as we believe that we are capable of adapting[Dwe12], we can be stable and calm in the face of adversity.

- Sai Krishna Rallabandi

Research Overview

Low Resource and Unwritten Languages

In order to bring technology closer to humans we believe it is important to provide mechanisms for them to interact with the same. For instance, language technologies such as Speech Recognition and Synthesis, Visual Question Answering to name a few have been shown to be massively useful in this context. However, these technologies are currently only accessible in a handful number of languages around the planet. In order to have a meaningful impact it is imperative that such technologies need to at least exist in many more languages. A major bottleneck in the progress of many data-intensive language processing tasks such as speech recognition and synthesis is scalability to new languages and domains. Building such technologies for unwritten or under-resourced languages is often not feasible due to lack of annotated data or other expensive resources. We are perhaps years away from having large enough datasets required to build robust models in these scenarios. Therefore, it might be interesting to investigate approaches that provide an alternate formulation.

A fundamental resource required to build such a stack is a phonetic lexicon - something that translates acoustic input to textual representation. Having such a lexicon, even if noisy, can help bootstrap speech recognition models, synthesis, and other technologies. We have been working towards developing datasets and models that generate such lexicons[Li+20a; Sch+18]. An example data contribution is the Wilderness speech processing dataset with over 650 language[Bla18]. An example model in this context is Allosaurus[Li+20a; Li+20b]. Allosaurus is a pretrained universal phone recognizer which can used to recognize phones in more than 2000 languages. In brief, this model attempts to incorporate phonology information into a multilingual speech recognition model through a special layer. The model first computes the phone distribution using an Automatic Speech Recognition(ASR) encoder and then the allophone layer maps the decoded phone distribution into the phoneme distribution for each language.

As an example application, we now present the work we are currently

doing in the space of Spoken Language Understanding (SLU). SLU is the task of processing spoken utterances to perform downstream tasks like Intent Recognition, Slot Filling etc. SLU systems are usually realized in the form of Spoken Dialogue Systems- Siri, Google Assistant and Alexa being the most popular examples. We have been working on building Spoken Dialogue Systems for low resource languages. The central idea behind our work is similar to the reasons behind building an E2E-SLU system - to explore the research question 'Do we really need to build full scale Speech to Text and Natural Language Understanding modules to realize spoken dialogue systems?' To deal with this problem, we have created a universal reduced representation of audio (in terms of phonemes when using Allosaurus[Li+20b]). The complexity of creating full-blown STT systems, which need to be built separately for different languages due to differences in written forms, is reduced to representing speech into phonemes, which are roughly universal across languages. On top of that, this technique also generalizes seamlessly to unwritten languages. These reduced representations of speech are then used to perform downstream tasks such intent recognition, slot filing etc. In [Gup+20], we do this for various Indian and European languages and show that our methods attain deployable performance for both language families.

Code Mixing

Code-switching (or mixing) refers to the phenomenon where bilingual speakers alternate between the languages while speaking. It occurs in every multilingual society such as India, Singapore, etc. It is used both to express opinions as well as for personal and group communications. This can go beyond simple borrowing of words from one language in another and is manifested at lexical, phrasal, grammatical and morphological levels. The technology today - from speech processing systems through conversational agents - assume monolingual mode of operation and do not process code-switched content. However, the mixed content is intuitively the most important part in the content. Since the systems are now handling conversations, it becomes important that they handle code-switching. Code switching is one of our highest conviction areas of research and we have been working

towards developing models that handle code switching in both speech and text modalities.

In the context of text to speech, there are two semi formal scenarios that are characterized by mixed data: (a) News paper headlines where the content is primarily in native language (say, Hindi) with English words interspersed and (b) Navigation instructions where the content is primarily in English with named entities in the native language. We have investigated several model variants towards these scenarios[Cha+17a; RB17]. We are currently interested in acoustic modeling of code mixed speech[GRB20] and speech recognition[Dal+20; RSB18]. In the context of text, perhaps one of the first applications where code mixing presents a deployment challenge is web search. Current information retrieval systems are only designed to work with a single interaction language. This assumption makes it inconvenient for multilingual users to interact naturally with the QA system. This problem gets aggravated in scenarios where they do not know the right word in the target language. To address this issue, we have built WebShodh - an end-end web-based Factoid QA system for CM languages. The system is demonstrated with two CM language pairs: Hinglish and Tenglish. Since it is unreasonable to assume existence of linguistic resources such as POS taggers or parsers for CM languages, WebShodh uses only bilingual dictionaries. More details about this can be found in [Cha+17b]. To nudge the community in this direction more actively, we have organized a challenge by systematically curating data in three CM languages - Hinglish, Tenglish and Tamlish. Details about the data collection and curation alongside an overview of participant systems can be found in [Cha+18a]. Further, we explicitly show the necessity to incorporate linguistic information in the context of code mixing:

- *Language Modeling*: Based on the hypothesis that encoding language information strengthens a the model, we empirically show the effectiveness of incorporating such information [Cha+18b].
- *Named Entity Recognition*: Exploiting the sparse nature of the token distribution, we have investigated approaches that can perform Named Entity Recognition from CM data [GCB18].
- *Style Categorization*: Along similar lines, we present a way to incorpo-

rate linguistic knowledge for recognizing CM content in [RSB18].

Natural Language Generation(NLG)

NLG is our highest conviction areas of research as of today. This century thus far has already witnessed a significant shift in the way we interact with technology and NLG has played a crucial role in this transition. In addition, NLG has a lot of applications in the form of data augmentation. We are investigating this research area in the context of unimodal text, unimodal speech as well as joint text and speech perspectives in the form of spoken dialog systems. Specifically, we are interested in controllable generation techniques[CNB19; PBS20]. You can find a neat organization of prior work[CB20] and a schema targeted at controllable NLG in [PBS20]. We present a brief overview of one such approaches here, being pioneered by Khyathi.

Anchoring Multimodal Generation

This year, I have approached the anchored generation from the groundwork that we laid last year. The idea is to support generation with content, structure and surface form realization. To this end, I worked on an anchoring framework that supports the improvement of these properties in 2 granularities of supervision: local and global. For each of the narrative properties, we demonstrate the utility of these levels of supervision.

The commonality across these properties is that these anchors are learnt from weak supervision using off-the-shelf tools. For instance, we use entity recognizers, finite state machines and lexical level language identification. With these tools and proposed models in place, the idea is to put together a large scaled equivalent of visual narratives for procedures, similar to 'how-to' descriptions and demonstrate the differences in techniques that we can leverage for each of these categories. So, we extend our previous work on gathering data for storyboarding recipes [CNB19] and introduce a new dataset and task of Visual Procedure Telling. We experiment with different techniques of infilling and replacement strategies to learn generating text from a sequence of images [CDB20]. But, is it possible that we always have a sequence of images to generate text for? What if there are no images

and we want to bridge the gap and make do with the set of few images we have where we want to still generate an elaborate and satisfying text. The most interesting takeaway for me from this work is the flexibility of the training procedure that invokes generating possibly missing contexts in real world scenarios with data imputation or missing sequences of images.

On a similar nerve, I have also worked on extending the concept of skeleton based generation [DCB19] to denoise content in large scaled automatically gathered datasets and better utilize them to improve captions in English and five other languages, including French, Italian, German, Spanish and Hindi [Cha+20]. As researchers prove time and again that our metrics are not good enough to capture the quality of generation, we also conducted large scale human evaluations to get relative and absolute scores for our captions and show that our skeleton based approach is better. Along with this, we also demonstrated the utility of these skeletons as an interpretable human-in-the-loop layer in controlling factors such as length, informativeness and gender debiasing of image captions.

Tools

Tools are an indispensable part of the academic research life. All of us employ different tools to allow us perform tasks more efficiently. In this part, we highlight a couple of tools which we believe have the potential to generate exponential returns over time with minimal start up cost.

Media Wiki - Xinjian Li

I am using media wiki to organize my daily thoughts and research logs, I find it very convenient to keep records and has been using it since 2015.

For those of you who are not familiar with media wiki, it is an open source software which is basically the software hosting Wikipedia! and most of the wiki projects. It has the same interface of Wikipedia, so most users should be familiar with browsing/searching/editing media wiki quite easily. Unlike the general wiki projects which expects many users to browse and edit, I find that media wiki is a perfect software to manage your personal contents. For example, I was using media wiki to keep my research logs, diary, photos, notes, and many other things.

There are a couple of good features media wiki could provide:

- First, it is free! you can freely install it on your own server and map your personal domain to it. For example, I mapped my sub domain https://wiki.xinjianl.com to it.
- Second, privacy is good. You can setup the password to prevent other people from accessing it. For example, you cannot see any of my contents from my wiki
- Third, backup is relatively easy, you can simply dump your contents from MySQL and stored it somewhere, I backup my wiki every month.
- Media wiki has many plugins so you can install them to enhance many features (syntax highlight, math rendering)
- Multimedia files are generally supported, you can upload audio/video/image to your wiki easily.
- There is an Emacs client so you can edit your wiki from your Emacs!!

However, there are also several cons:
- You need to regularly update it to make sure it is secure
- Some plugins are not maintained anymore
- Your server will cost some money

Despite these drawbacks, I think it is one of the best options to manage your own contents!

Similar to mediawiki are tools like Notion and Roam Research. Here's what Peter Wu has to say about Notion:

> *I really like Notion because of its support for a few great features: tables, keyboard shortcuts, note organization, nice formatting e.g. dividers and bullets. I've been mostly using the table feature because it supports sorting and filtering. Additionally, I like how Notion lets us link notes fairly well, though I haven't taken full advantage of this feature.*

Neo4j - Sai Krishna Rallabandi

I have been using the graph database Neo4j for literature review from a couple of years now. Using Neo4j, I semi automatically annotate each research entry - be it a tweet or a complete research paper with two labels: (a) code source or (b) literature source. These entries are encoded as nodes and are matched to what I refer to as super nodes. For instance, the entry 'Attention is not not an Explanation', which is a research paper, is encoded as a node while 'Attention' is encoded as a super node. I have a routine that periodically populates nodes crawling literature from various sources including ArXiv and Google scholar and posts from reddit, twitter to help update the database. Neo4j provides research topic visualization employing this database as the back end.

Here are some of the pros:
- The UI is very intuitive and makes it easy to visualize.

- Neo4j has its own query language CYPHER which is extremely intuitive. Inserting or updating any node or relation is also very easy both through the UI or a script.
- The interface is efficient on large datasets to support the multiple relations between the nodes. This perhaps is the reason many companies employ Neo4j internally as well.
- It is also extremely fast. Therefore, the tool comes handy when we dont know the schema beforehand and we are looking to identify schema/pattern. For instance, I was able to identify multiple 'trends' from Interspeech 2020[7] since I was able to quickly sift through various query types.
- It has built in web interface and there are a lot of third party libraries supporting a variety of functionality.

Here are some features I wish they implemented:

- Since I love the concept of name spaces, it would be great to have something similar.
- Memory allocation could be improved.
- At this moment, text normalization has to be performed manually. For instance, co-attention and Co Attention and co-attn register as different nodes in the graph. It would be great if there is a mechanism to automate this within the tool.

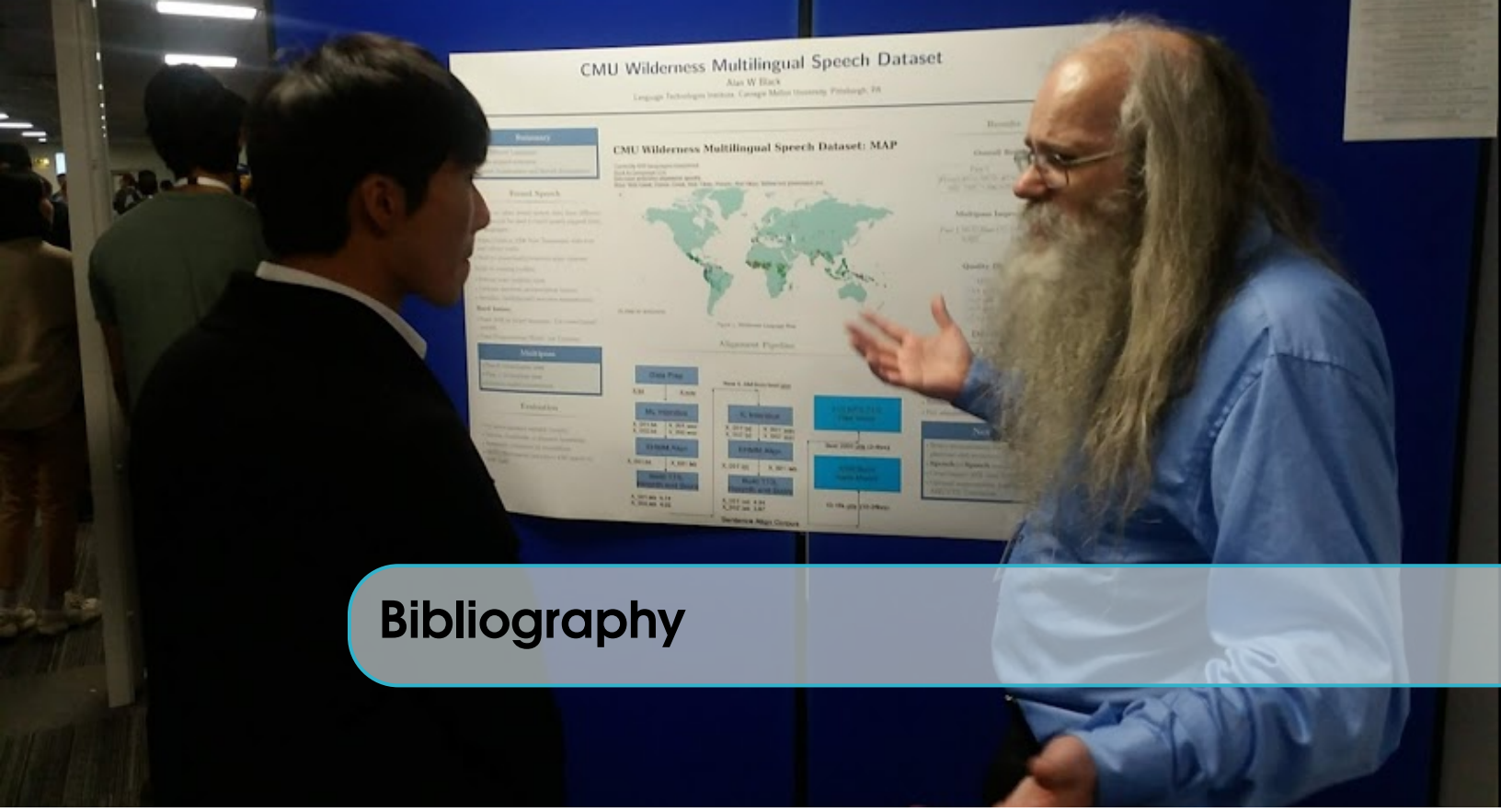
Overall, I am very impressed with the tool itself and it has helped me visualize topics within my dissertation very satisfactorily.

[7] t.ly/kHrN

Computational Resources

The table below summarizes our computational resources in terms of GPU. Although we anticipated a lockdown and bought GPUs March this year, we have not been able to assemble the machine in time to set up a new one in 2020. We should be back to building newer machines thereby increasing per student GPU memory quota in 2021. K13 is the only machine with a Tesla K40 GPU. Rest of them are either 1080 Ti or 1060.

Year	**n(machines)**	**Memory(Gigs)**	**Machines**
2017	1	12	K13
2018	2	36	K13,K20
2019	4	88	K13,K20,K21,K22
2020	3	76	K20,K21,K22



Bibliography

[Ara11] Cynthia Aranow. "Vitamin D and the immune system". In: *Journal of investigative medicine* 59.6 (2011), pages 881–886 (cited on page 8).

[AWM11] James B Avey, Tara S Wernsing, and Ketan H Mhatre. "A longitudinal analysis of positive psychological constructs and emotions on stress, anxiety, and well-being". In: *Journal of Leadership & Organizational Studies* 18.2 (2011), pages 216–228 (cited on page 6).

[Awu+98] Emmanuel MK Awumey et al. "Vitamin D metabolism is altered in Asian Indians in the southern United States: a clinical research center study". In: *The Journal of Clinical Endocrinology & Metabolism* 83.1 (1998), pages 169–173 (cited on page 8).

[Bae+10] Femke Baeke et al. "Vitamin D: modulator of the immune system". In: *Current opinion in pharmacology* 10.4 (2010), pages 482–496 (cited on page 8).

[Bla18] Alan W Black. "Wilderness Dataset". In: *https://github.com/festvox/datasets-CMUWilderness* (2018) (cited on page 11).

[BMR11] Maria C Borges, Lıgia A Martini, and Marcelo M Rogero. "Current perspectives on vitamin D, immune system, and chronic diseases". In: *Nutrition* 27.4 (2011), pages 399–404 (cited on page 8).

[BAW11] Richard A Burns, Kaarin J Anstey, and Timothy D Windsor. "Subjective well-being mediates the effects of resilience and mastery on depression and anxiety in a large community sample of young and middle-aged adults". In: *Australian & New Zealand Journal of Psychiatry* 45.3 (2011), pages 240–248 (cited on page 6).

[BPB10] Nicola W Burton, Ken I Pakenham, and Wendy J Brown. "Feasibility and effectiveness of psychosocial resilience training: a pilot study of the READY program". In: *Psychology, health & medicine* 15.3 (2010), pages 266–277 (cited on page 6).

[CNB19] Khyathi Chandu, Eric Nyberg, and Alan W Black. "Storyboarding of Recipes: Grounded Contextual Generation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, pages 6040–6046 (cited on page 14).

[Cha+18a] Khyathi Chandu et al. "Code-Mixed Question Answering Challenge: Crowd-sourcing Data and Techniques". In: *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. 2018, pages 29–38 (cited on page 13).

[Cha+18b] Khyathi Chandu et al. “Language informed modeling of code-switched text”. In: *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. 2018, pages 92–97 (cited on page 13).

[CB20] Khyathi Raghavi Chandu and Alan W Black. “Dissecting the components and factors of Neural Text Generation”. In: *arXiv preprint arXiv:2010.07279* (2020) (cited on page 14).

[CDB20] Khyathi Raghavi Chandu, Ruo-Ping Dong, and Alan W Black. “Reading between the Lines: Exploring Infilling in Visual Narratives”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pages 1220–1229 (cited on page 14).

[Cha+17a] Khyathi Raghavi Chandu et al. “Speech Synthesis for Mixed-Language Navigation Instructions”. In: *Proc. Interspeech 2017* (2017) (cited on page 13).

[Cha+17b] Khyathi Raghavi Chandu et al. “WebShodh: A Code Mixed Factoid Question Answering System for Web”. In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer. 2017, pages 104–111 (cited on page 13).

[Cha+20] Khyathi Raghavi Chandu et al. “Weakly Supervised Content Selection for Improved Image Captioning”. In: *CoRR* abs/2009.05175 (2020). arXiv: 2009.05175. URL: https://arxiv.org/abs/2009.05175 (cited on page 15).

[Dal+20] Siddharth Dalmia et al. “Transformer-Transducers for Code-Switched Speech Recognition”. In: *arXiv preprint arXiv:2011.15023* (2020) (cited on page 13).

[Dem+19] Taye Demeke et al. “Comorbidity and health-related quality of life in Somali women living in Sweden”. In: *Scandinavian Journal of Primary Health Care* 37.2 (2019), pages 174–181 (cited on page 7).

[DCB19] Ruo-Ping Dong, Khyathi Raghavi Chandu, and Alan W Black. “Induction and Reference of Entities in a Visual Story”. In: *arXiv preprint arXiv:1909.09699* (2019) (cited on page 15).

[Dwe12] Carol Dweck. *Mindset: Changing the way you think to fulfil your potential*. Hachette UK, 2012 (cited on page 10).

[Epe+09] Elissa Epel et al. “Can meditation slow rate of cellular aging? Cognitive stress, mindfulness, and telomeres”. In: *Annals of the new York Academy of Sciences* 1172 (2009), page 34 (cited on page 9).

[GCB18] Parvathy Geetha, Khyathi Chandu, and Alan W Black. “Tackling Code-Switched NER: Participation of CMU”. In: *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. 2018, pages 126–131 (cited on page 13).

[Gra+16] Simona Graziano et al. “Vitamin D/vitamin D receptor axis regulates DNA repair during oncogene-induced senescence”. In: *Oncogene* 35.41 (2016), pages 5362–5376 (cited on page 7).

[Guc+11] Daniel F Gucciardi et al. “The Connor-Davidson Resilience Scale (CD-RISC): Dimensionality and age-related measurement invariance with Australian cricketers”. In: *Psychology of Sport and Exercise* 12.4 (2011), pages 423–433 (cited on page 6).

[GRB20] Akshat Gupta, Sai Krishna Rallabandi, and Alan W Black. “Mere account mein kitna balance hai?–On building voice enabled Banking Services for Multilingual Communities”. In: *arXiv preprint arXiv:2010.16411* (2020) (cited on page 13).

[Gup+20] Akshat Gupta et al. “Acoustics Based Intent Recognition Using Discovered Phonetic Units for Low Resource Languages”. In: *arXiv preprint arXiv:2011.03646* (2020) (cited on page 12).

[HRM76] Gershon W Hepner, Martin Roginsky, and Hing Fai Moo. “Abnormal vitamin D metabolism in patients with cirrhosis”. In: *The American journal of digestive diseases* 21.7 (1976), pages 527–532 (cited on page 7).

[Hew12] Martin Hewison. “Vitamin D and the immune system: new perspectives on an old theme”. In: *Rheumatic Disease Clinics* 38.1 (2012), pages 125–139 (cited on page 8).

[HSH13] Arash Hossein-Nezhad, Avrum Spira, and Michael F Holick. “Influence of vitamin D status and vitamin D 3 supplementation on genome wide expression of white blood cells: a randomized double-blind clinical trial”. In: *PloS one* 8.3 (2013), e58725 (cited on page 7).

[HZW15] Tianqiang Hu, Dajun Zhang, and Jinliang Wang. “A meta-analysis of the trait resilience and mental health”. In: *Personality and Individual Differences* 76 (2015), pages 18–27 (cited on page 6).

[Ji+10] Li Li Ji et al. “Exercise-induced hormesis may help healthy aging”. In: *Dose-Response* 8.1 (2010), dose–response (cited on page 9).

[Li+20a] Xinjian Li et al. “Towards Zero-Shot Learning for Automatic Phonemic Transcription.” In: *AAAI*. 2020, pages 8261–8268 (cited on page 11).

[Li+20b] Xinjian Li et al. “Universal phone recognition with a multilingual allophone system”. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pages 8249–8253 (cited on pages 11, 12).

[LJ18] Paul Lips and Renate T de Jongh. “Vitamin D deficiency in immigrants”. In: *Bone reports* 9 (2018), pages 37–41 (cited on page 8).

[Mas+11] Caitlin Mason et al. “Effects of weight loss on serum vitamin D in postmenopausal women”. In: *The American journal of clinical nutrition* 94.1 (2011), pages 95–103 (cited on page 7).

[Mee+11] Irene M van der Meer et al. “Prevalence of vitamin D deficiency among Turkish, Moroccan, Indian and sub-Sahara African populations in Europe and their countries of origin: an overview”. In: *Osteoporosis international* 22.4 (2011), pages 1009–1021 (cited on page 8).

[Meh+19] Mansi H Mehta et al. “Examining the positive cognitive triad: a link between resilience and well-being”. In: *Psychological reports* 122.3 (2019), pages 776–788 (cited on page 6).

[Min+13] Jung-Ah Min et al. “Psychological resilience contributes to low emotional distress in cancer patients”. In: *Supportive care in cancer* 21.9 (2013), pages 2469–2476 (cited on page 6).

[Mor05] Howard A Morris. “Vitamin D: a hormone for all seasons-how much is enough? Understanding the new pressures”. In: *Clinical Biochemist Reviews* 26.1 (2005), page 21 (cited on page 7).

[NSC19] Veijo Nurminen, Sabine Seuter, and Carsten Carlberg. “Primary vitamin D target genes of human monocytes”. In: *Frontiers in physiology* 10 (2019), page 194 (cited on page 7).

[PMD98] Gerald D Parr, Marilyn Montgomery, and Camille DeBell. “Flow theory as a model for enhancing student resilience”. In: *Professional School Counseling* 1.5 (1998), pages 26–31 (cited on page 6).

[Par98] Timo Partonen. “Vitamin D and serotonin in winter”. In: *Medical hypotheses* 51.3 (1998), pages 267–268 (cited on page 8).

[PBS20] Shrimai Prabhumoye, Alan W Black, and Ruslan Salakhutdinov. “Exploring Controllable Text Generation Techniques”. In: *arXiv preprint arXiv:2005.01822* (2020) (cited on page 14).

[RB17] SaiKrishna Rallabandi and Alan W Black. “On building mixed lingual speech synthesis systems”. In: *Proc. Interspeech 2017* (2017) (cited on page 13).

[RSB18] SaiKrishna Rallabandi, Sunayana Sitaram, and Alan W Black. “Automatic Detection of Code-switching Style from Acoustics”. In: *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. 2018, pages 76–81 (cited on pages 13, 14).

[Rob+15] Ivan T Robertson et al. “Resilience training in the workplace from 2003 to 2014: A systematic review”. In: *Journal of occupational and organizational psychology* 88.3 (2015), pages 533–562 (cited on page 6).

[Sch+18] Odette Scharenborg et al. “Linguistic unit discovery from multi-modal inputs in unwritten languages: Summary of the “Speaking rosetta” JSALT 2017 workshop”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pages 4979–4983 (cited on page 11).

[SPB20] Amrith Setlur, Barnabás Póczos, and Alan W Black. “Nonlinear ISA with Auxiliary Variables for Learning Speech Representations”. In: *arXiv preprint arXiv:2007.12948* (2020) (cited on page 4).

[Soo+11] Amit Sood et al. “Stress management and resilience training among Department of Medicine faculty: a pilot randomized clinical trial”. In: *Journal of general internal medicine* 26.8 (2011), pages 858–861 (cited on page 6).

[Tar+17] Adrienne A Taren et al. “Mindfulness meditation training and executive control network resting state functional connectivity: a randomized controlled trial”. In: *Psychosomatic medicine* 79.6 (2017), page 674 (cited on page 9).

[Ves+11] Peter Vestergaard et al. “Vitamin D status and bone and connective tissue turnover in brown bears (Ursus arctos) during hibernation and the active state”. In: *PLoS One* 6.6 (2011), e21483 (cited on page 7).

[Wam+13] L Wamberg et al. “Expression of vitamin D-metabolizing enzymes in human adipose tissue—the effect of obesity and diet-induced weight loss”. In: *International journal of obesity* 37.5 (2013), pages 651–657 (cited on page 7).

[WKH88] Ann R Webb, Loren Kline, and Michael F Holick. “Influence of season and latitude on the cutaneous synthesis of vitamin D3: exposure to winter sunlight in Boston and Edmonton will not promote vitamin D3 synthesis in human skin”. In: *The journal of clinical endocrinology & metabolism* 67.2 (1988), pages 373–378 (cited on page 7).