

Notas NeurIPS 2019

André Pacheco
pacheco.comp@gmail.com

Vancouver, BC, Canada
Dezembro de 2019

Conteúdo

Notas NeurIPS 2019	1
1 Introdução	2
1.1 Sobre a conferência	3
1.2 Oportunidades e vantagens de participar NeurIPS	4
1.3 Organização deste documento	4
2 Dia 1: Sponsor Expo	4
2.1 Workshop - <i>Multi-modal Research To Production</i>	5
2.2 Palestras	5
Google: Interpretability - Now What?	5
Apple: Private Federated Learning doc.ai: Federated Learning in Healthcare	7
Unity: how the gaming industry is driving in AI research Sony: Computational creativity for Music and Gastronomy applications	8
3 Dia 2: Tutoriais e workshops	8
3.1 3 Systems perspectives into Human-Centered Machine Learning - Carlos Guestrin	8
3.2 Machine learning for computational biology and health - Anna Goldenberg and Barbara Engelhardt	11
3.3 How to know - Celeste Kidd	13
3.4 Mais fontes interessantes	15
4 Dia 3: Apresentações orais e pôsteres I	15
4.1 Apresentações orais	16
Deep equilibrium models	16
Towards explaining the regularization effect of initial large learning rate in training neural networks	17
4.2 Pôsteres	18
5 Dia 4: Apresentações orais, pôsteres II + evento satélite UBC	19
5.1 Machine Learning in Medical Imaging : Progress So Far and Perspectives - Thad Hughes	19

5.2	Deep learning for intelligent medical imaging solutions - Julia Schnabel	21
5.3	AI in healthcare working towards positive clinical impact - Nenad Tomašev	22
5.4	From System 1 Deep Learning to System 2 Deep Learning - Yoshua Bengio	22
	Lidando com mudanças na distribuição	25
	Blocos básicos para o sistema 2: <i>attention & consciousness</i>	26
	Consciousness prior: sparse factor graph	27
	Framework teórico	27
	Composição de arquiteturas em DL	27
5.5	Pôsteres	27
5.6	Mais fontes interessantes	28
6	Dia 5: Apresentações orais, pôsteres III	28
6.1	Agency + Automation: Designing Artificial Intelligence into Interactive Systems - Jeffrey Heer	28
	Exploração e visualização de dados	29
	<i>Data cleaning e Transformation</i>	30
	<i>Natural Language translation</i>	30
	Desafios futuros	30
6.2	Apresentações orais	31
	Adversarial Music: Real World Audio Adversary Against Wake-word Detection System	31
	This Looks Like That: Deep Learning for Interpretable Image Recognition	31
6.3	Pôsteres	31
6.4	Mais fontes interessantes	33
7	Dia 6 e 7: workshops	33
	Palestras	34
7.1	Pôsteres	35
8	Considerações finais	36

1 Introdução

Este ano eu tive a oportunidade de participar da [NeurIPS 2019](#) que aconteceu em Vancouver entre os dias 8 e 14 de dezembro. O principal motivo foi para apresentar um [artigo](#) em um dos *workshops* da conferência. Porém, eu estive presente em todos os dias do evento e neste documento pretendo compartilhar as tendências e *insights* das palestras, apresentações e painéis que eu consegui assistir. Friso a palavra **consegui** porque a conferência é **enorme**. São dezenas de eventos em paralelo ocorrendo no local do evento e fora. Por exemplo, a *University of British Columbia* (UBC) ofereceu um evento satélite em um dos dias da conferência na qual eu participei e vou citar mais adiante. Por isso, quero começar descrevendo os números e características da conferência, oportunidades

e vantagens de participar da mesma.

Caso você encontre algum erro ou algo que ache interessante incluir aqui, não hesite em entrar em contato via email: pacheco.comp@gmail.com.

1.1 Sobre a conferência

Como disse, a NeurIPS é uma conferência enorme. Na abertura oficial eles apresentaram números que estão detalhados em um [post no blog da conferência no medium](#). Você pode encontrar todos os detalhes no post, aqui quero destacar apenas alguns deles:

- **13k pessoas** se registraram para participar da conferência.
- 6743 artigos foram submetidos e 1428 foram aceitos sendo 1228 como posters.
- A taxa de aceite da conferência foi de **21,6%**.
- A % de aceite por área está ilustrado na Figura 1. Neste ano, *neuroscience* e *theory* representam mais de 60% dos artigos.
- 51 *workshops* foram aceitos na conferências sendo 4 para incluir representatividade e minorias: LatinX in AI, Black in AI, Queer in AI e Women in AI.
- Contando todos os artigos aceitos, são mais de **16k páginas**

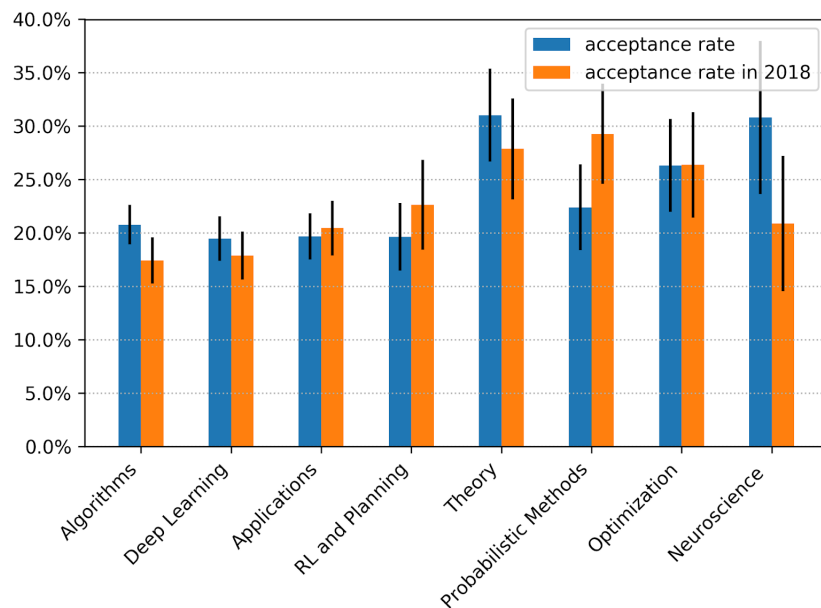


Figura 1. % de artigos aceitos por área de conhecimento ([fonte](#))

Então, como pode perceber, a conferência é gigante e é humanamente impossível digerir tudo isso. A parte boa é que neste ano tudo está filmado e disponível gratuitamente [neste link](#). Assim, você pode assistir algo que lhe interessa. No fim desse documento recomendo algumas palestras, painéis e apresentações de papers. O cronograma oficial da conferência você encontra [aqui](#).

1.2 Oportunidades e vantagens de participar NeurIPS

A NeurIPS é a maior conferência da área de *machine learning*. Participar da conferência é uma oportunidade única de conversar e interagir com pesquisadores de grandes universidades e empresas como Google, OpenAI, Facebook, etc. Dentro da conferência existem diversos *meet-ups* e *sociais* que existem para que você troque ideias e possa interagir com outros pesquisadores da sua área. Porém, sabemos que nós brasileiros não temos tantos recursos para pagar inscrição, viajar, hospedar etc. Por isso é importante checar o *travel awards* da conferência. Caso você tenha um artigo aceito é muito mais fácil de conseguir. Esse ano foram distribuídos 400 e eu fui um dos contemplados. No caso, eles cobriram a minha inscrição e me ajudaram com U\$ 500,00 para passagem e hospedagem. Com esse valor eu consegui pagar a passagem aérea apenas. A hospedagem foi na casa de um amigo (olha a importância do *networking*) e alimentação por minha conta. De qualquer forma, o auxílio foi fundamental.

Por fim, gostaria de destacar os *workshops*. Na NeurIPS é quase como uma sessão especial sobre algum assunto ou sobre um determinado grupo. O [LatinX in AI](#), por exemplo, é um grupo voltado para nossa comunidade latina. Eles também oferecem *travel awards* por fora dos 400 da conferência. Logo, é uma excelente oportunidade caso você queira publicar em 2020. De maneira geral, os *workshops* é uma alternativa mais acessível a conferência e caso você tenha interesse de participar da mesma em 2020 eu recomendo que fique de olho neles.

1.3 Organização deste documento

Como são 7 dias de conferência e eu estava em todos, esse documento está organizado por ordem cronológica. Logo, cada seção começará com o dia e o principal evento que ele representa. Dessa maneira, acho mais fácil organizar minhas ideias e minhas anotações, uma vez que boa parte desse documento foi escrito/editado ao longo da conferência.

2 Dia 1: Sponsor Expo

O primeiro dia da conferência é destinado aos patrocinadores, como Google, Intel, Apple, Facebook, Microsoft, Sony e muitos outros. As empresas ofereceram mini-cursos, palestras, demonstrações e paines sobre pesquisa e novas tecnologias. Novamente, é impossível frequentar tudo pois era simultâneo. A seguir eu descrevo minhas principais anotações sobre esse dia.

2.1 Workshop - *Multi-modal Research To Production*

Esse *workshop* foi oferecido pelo Facebook e tinha como objetivo difundir o uso de suas ferramentas, no caso o PyTorch, e mostrar novas tecnologias que eles estão desenvolvendo. Eu participei de todo *workshop* pois durava 4h e era na mesma hora que algumas palestras que eu estava mais interessado. De maneira geral, eles apresentaram:

- Conceitos básicos do PyTorch para quem não conhecia a linguagem
- Como eles escalam PyTorch para produção. Neste caso, eles usam o *torchscript*, que transcreve o código para C++ para aumentar o desempenho computacional. Um tutorial para essa ferramenta pode ser encontrada [aqui](#).
- Apresentaram resultados da pesquisa deles em relação a *Quantization*, uma técnica para aumentar a performance computacional e armazenar tensores com precisão menor do que float. Um artigo discutindo essa ideia pode ser encontrado [aqui](#).
- O mais novo *framework open-source* que eles desenvolveram chamado *Classy Vision*, que nada mais é do que uma ferramenta para treinar modelos de maneira mais fácil e rápida. Na minha opinião ela é parecida com o Keras e é mais voltada para empresas do que pesquisa. Mas, é bastante útil e rápido. Para mais informações, acesse o [GitHub](#) deles.

Eu consegui conversar com alguns desenvolvedores que trabalham na equipe do PyTorch e eles estão muito interessados em ver a ferramenta deles funcionando em projetos que envolvam a produção. Inclusive, pediram para compartilhar com eles exemplos de modelos embarcados, como por exemplo em smartphones. Eles estão bem focados nessa parte e trabalhando para melhorar ela cada vez mais.

2.2 Palestras

As palestras eram eventos maiores que as empresas tinham 25 min para apresentar um projeto de pesquisa ou produto que eles estão desenvolvendo. Algumas apresentaram tópicos semelhantes e eu juntei na mesma subseção aqui.

Google: Interpretability - Now What? Essa palestra foi sobre como interpretar o resultado de um *deep model*. Atualmente, *deep learning* funciona como uma *black-box*. O modelo oferece o resultado e temos que confiar nele. Porém, como interpretar esses resultados? Será que podemos confiar nele? Esse tema certamente foi uma das tendências da conferência e várias palestras e apresentações investiram neste assunto.

O objetivo da interpretabilidade é deixar o uso de *machine learning* mais responsável. Para isso, é necessário garantir:

1. Que nossos valores estejam alinhados com o modelo
2. Que nosso conhecimento seja refletido para todos

Por outro lado, não é um objetivo da interpretabilidade:

1. Fazer com que todos os modelos sejam interpretáveis
2. Entender absolutamente tudo sobre o modelo
3. Apenas aumentar a confiança do usuário no modelo

A principal pergunta aqui é: dado a predição para uma determinada classe, qual é a **evidência** dessa predição? Um dos métodos mais famosos é conhecido como [Saliency maps](#), que tenta medir a importância de cada pixel na imagem. Porém, mesmo com uma predição aleatória, o método não muda tanto o conjunto de pixels mais importantes, como foi mostrado [neste paper](#). Uma conclusão importante é que não apenas por fazer sentido para um humano que a evidência da predição está correta.

Para elevar a dificuldade dos testes de interpretabilidade eles propuseram o [Benchmarking interpretability methods \(BIM\)](#), um conjunto de imagens como a ilustrada na Figura 2. Para o caso dessa imagem em específico, na predição de uma floresta o cachorro não pode fazer parte da evidência que explica a predição.



Figura 2. Exemplos de imagens do BIM ([fonte](#))

Um método que propuseram para lidar com esse problema foi o [Testing with Concept Activation Vectors \(TCAV\)](#). A ideia é uma explicação quantitativa: o quanto de um conceito foi importante para uma dada predição mesmo que esse conceito não faça parte do treinamento. Para definir esses conceitos é necessário exemplos, e este, na minha opinião é um ponto fraco do método. Um exemplo dado durante a palestra: listas pretas e brancas é um conceito importante para identificação de uma zebra? Esse conceito precisa ser conhecido previamente e ter exemplos do mesmo.

Para concluir, foi mostrado que o TCAV já vem sendo utilizado para problemas do mundo real, como predição de tempestades e auxiliando doutores a entender predição de doenças. Além disso, foi listado alguns *takeaways*:

- É necessário ter boas avaliações para o problema
- Testar a interpretabilidade com humanos é bom, mas tem que ter cuidado com bias e irracionalidade
- Ser crítico com a explicação e sempre checar se a interpretação está indo para a direção correta

Apple: Private Federated Learning | doc.ai: Federated Learning in Healthcare Essa duas palestras tiveram o mesmo tema, logo eu as juntei na mesma subseção. *Federated learning* foi outra tendência da conferência. De maneira geral é um conceito novo em *machine learning* que visa lidar, principalmente, com a privacidade do usuário. A ideia é relativamente simples e está ilustrada na Figura 3. Imagine que você possui um sistema que coleta dados de usuários via smartphone. Ao invés de enviar esses dados para um servidor e treinar um modelo para um determinado fim, você roda modelos menores no smartphone do usuário e envia apenas os pesos para servidor. Ao final, os pesos de todos os modelos são agregados por uma simples média aritmética.

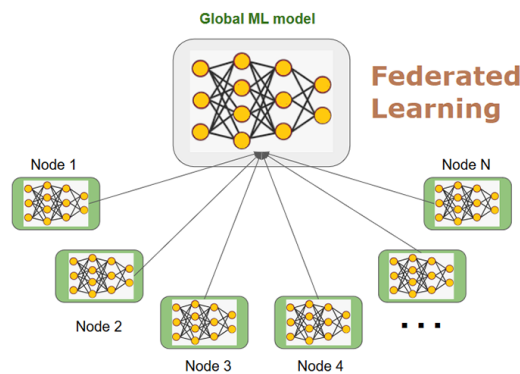


Figura 3. Exemplo do funcionamento do Federated learning ([fonte](#))

Periodicamente esse o modelo global é atualizado e enviado para os modelos locais, que continuam a treinar com a entrada de dados e retoma o loop anterior de atualizar o modelo global e assim por diante. Obviamente existem diversas limitações técnicas para implantar esse modelo, como limitação de recurso do smartphone. Além disso, na palestra da Apple eles chamaram a atenção que é possível recuperar os dados do usuário via ataque no peso das redes. Por conta disso, eles utilizam um filtro que realiza uma espécie de *soft* criptografia nos pesos para que isso não aconteça.

Na palestra do doc.ai eles mostraram um modelo deles que é treinado a partir de selfies do usuário. Eles mostraram resultados de como o método funciona mas

ainda não possuem um teoria que explica o porque. De qualquer forma essa é uma área nova e promissora que acredito que ainda vamos ouvir falar muito.

Unity: how the gaming industry is driving in AI research | Sony: Computational creativity for Music and Gastronomy applications Eu juntei essas duas palestras pois elas foram bem interessantes mas bem no âmbito de aplicações tecnológicas que estão sendo feitas com ML. Começando pela Unity, a palestra focou em como a indústria de jogos estão fomentando o avanço em *deep reinforcement learning*. Eles apresentaram um *framework open-source* que eles desenvolveram chamado [ML agents](#) que nada mais é do que uma *toolkit* para aplicar *deep learning* usando Unity. Aliás, essa é outra área super em alta na conferência e na área muito por conta da OpenAI 5 (o bot de Dota 2 que vou comentar mais para frente) e a Alpha X (o bot de StarCraft). Além dessa toolkit, a empresa está fomentando outras competições na conferência como a [Animal-AI Olympics](#) e [Obstacle Tower challenge](#). Caso você queira aprender e aplicar *deep reinforcement learning*, esse é um bom caminho.

Em relação a Sony, a empresa apresentou a AI que eles desenvolveram para composição de músicas. Eles fizeram uma demonstração de uma música composta a partir de um modelo que aprendeu com as músicas dos Beatles chamada [Daddy's car](#), que na minha opinião ficou bem legal de se ouvir. Eles fazem questão de deixar claro que essa ferramenta visa auxiliar os músicos e não substituí-los. De maneira geral, eles dizem dominar toda teoria e softwares para compor uma música é complexo, logo essa tecnologia é uma maneira de democratizar a música.

Por fim, eles apresentaram um conceito que estão trabalhando mas que ainda está muito no início. A ideia é desenvolver uma bancada que auxilia na preparação da comida. Eles demonstraram alguns vídeos de robôs sendo treinados e uma *concept art* que você pode [assistir aqui](#).

3 Dia 2: Tutoriais e workshops

O segundo dia de evento foi o dia dos tutoriais. Porém, também ocorreu o *workshop* do LatinX in AI, na qual eu participei, a cerimônia de abertura e a primeira palestra principal. A seguir descrevo minhas principais observações.

3.1 3 Systems perspectives into Human-Centered Machine Learning - Carlos Guestrin

Essa foi a palestra de abertura do *workshop* do LatinX. Na minha opinião, a melhor do evento. Você pode assistí-la [neste link](#).

Prof. Guestrin inicia a palestra fazendo uma breve revisão sobre o *momentum* em relação a ML e já inicia dizendo que **não existe ML sem a contribuição humana no processo**. Por isso ele foca a palestra em 3 áreas:

1. **Create:** como fazer com que mais pessoas criem modelos em ML

2. **Deploy:** como fazer deploy desses modelos
3. **Trust:** como fazer com que as pessoas entendam o que um modelo está prevendo adquirindo assim confiança no mesmo

Em relação a criação, ele apresenta brevemente ferramentas como XGBoost, Turi Create e Create ML, que segundo ele está democratizando o uso de ML.

Em seguida, ele parte para o segundo tópico que é sobre como colocar os modelos em produção. Ele usa um exemplo de que na academia basicamente cria-se um modelo, mostra que o resultado é melhor do que outro e publica-se um artigo. Porém, para fazer produção é bem mais complicado do que isso. É necessário realizar testes profundos, entender o device que o modelo vai rodar, etc. Neste ponto ele apresenta a CoreML, uma ferramenta que a Apple (empresa que ele trabalha) disponibiliza para desenvolvedores para acelerar o processamento dos modelos em ML. Aqui ele aproveita o gancho para citar um trabalho de um aluno dele em otimizar modelos de ML para aumentar o desempenho em produção. Esse trabalho pode ser acessado [aqui](#). Essa é uma área que ele aponta como desafiadora e cheia de oportunidades na área.

Por fim ele falar sobre a confiança na predição dos modelos. Neste caso, é basicamente a mesma questão de interpretabilidade da palestra do Google. Neste caso ele apresenta um método que seu grupo desenvolveu conhecido como [LIME](#) (curiosamente, eu estou utilizando esse método). Ele introduz um exemplo interessante. Eles treinaram um modelo para diferenciar lobos de cães da raça Husky. O modelo apresenta acurácia de 95%. Ao investigar as evidências que o modelo estava observando para tomar a decisão eles perceberam que quando havia neve na imagem, ele previa como lobo. Logo, eles treinaram um modelo profundo pra detectar neve. Um caso clássico de bias que está ilustrado na Figura 4

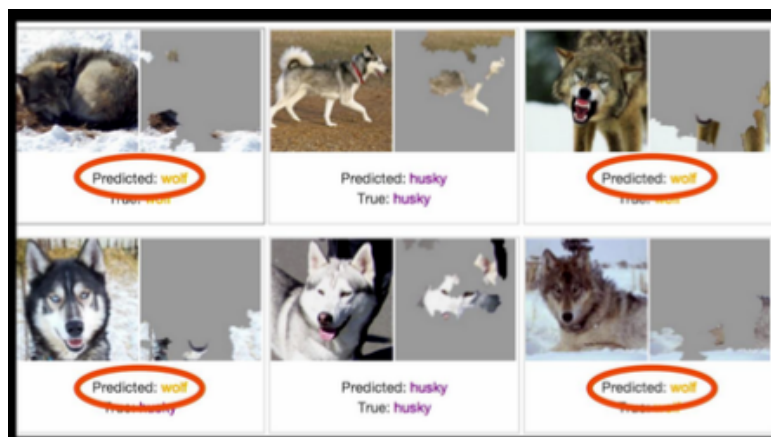


Figura 4. Partes da imagem que explicam a predição do modelo ([fonte](#))

Na visão dele, interpretabilidade em ML é: **fornecer a humanos um modelo mental do comportamento do modelo**. Porém, existe um trade-off entre interpretabilidade e acurácia. Modelos mais complexos com melhores acurácias são muito mais complexos de ser interpretados. Com isso, o LIME tenta aprender decisões locais ao invés de globais. Segundo ele, próximo da fronteira de decisão a predição se aproxima de um modelo linear.

Outro problema interessante em relação a robustez e interpretabilidade que eles estão trabalhando é no problema de [Visual Question and Answering](#). Uma das imagens mais famosas deste problema é mostra da Figura 5



Figura 5. Exemplo de uma imagem do problema de VQA ([fonte](#))

Para esta imagem são realizadas algumas perguntas:

1. What is the mustache made of? A: Banana
2. How many bananas are in the picture? A: 2

Analisando as evidências do modelo eles perceberam que as âncoras para essas duas perguntas são as palavras **what** e **how many**. Dessa forma, **qualquer pergunta** utilizando *what* retorna a resposta banana e *how many* retorna 2. Ele até brinca que se perguntar para o modelo: **What** is the head of the US? A resposta é banana.

Por fim, ele termina falando sobre ataques adversariais, não apenas para imagens, mas também para problemas de processamento de linguagem natural (NLP). Por exemplo, trocando palavras com mesmo significado semântico as respostas mudam completamente. Por exemplo, color por colour, what por which e assim por diante. Ele reporta esses problemas [neste paper](#).

Para concluir, ele fala mais um pouco sobre o problema de bias. Mostra exemplos de softwares racistas e misóginos. Esses problemas apresentados são muito importantes principalmente na área médica. E a comunidade de ML/AI deve tomar muito cuidado com isso por que as decisões e modelos criados com ML/AI impactam a vida de bilhões de pessoas. Por conta disso, atualmente existe um *momentum* em relação a teoria e técnicas em relação a ética da inteligência artificial.

3.2 Machine learning for computational biology and health - Anna Goldenberg and Barbara Engelhardt

Este foi um dos tutoriais do dia. Ele teve duração de 2h e foi basicamente um grande *survey* da área. O interessante é que as professoras iniciaram desde aplicações para células até *healthcare*. A apresentação completa está disponível [neste link](#).

Já como primeira lição, elas deixam claro que **aplicações de ML para biomedicina devem ser realizadas com profissionais da área**. Parece óbvio, mas a verdade é que a maioria não tem contribuição da área médica. É importantíssimo aprender sobre os dados e o domínio que se está desenvolvendo.

E a área de biomedicina é muito difícil para ML por alguns motivos:

- Existem muito mais features do que número de amostras
- Os dados possuem muito ruídos e caudas longas
- Supor que uma característica rara não ocorre é falho
- Muitas amostras não são similares a outras
- Dados são complexos e *high-dimensional*
- Sabe-se muito pouco sobre os mecanismos da biomedicina

E o mais importante: o impacto de solucionar problemas da biomedicina com ML afeta todos nós. Por isso temos que ter cautela sempre.

As grandes questões em relação a biomedicina hoje envolve o mecanismo biológico, ou seja, entender como os passos bioquímicos ocorre para um determinado evento (*mechanistic pathway*). Algumas dessas perguntas:

- Como a variação de população surge? É possível quantificá-la?
- Quais são os componentes essenciais e não essenciais do *mechanistic pathway*?
- Como ele funciona e por que ele é silencioso?

Essas perguntas são importantes para desenvolvimentos de remédios, tratamento e prevenção de doenças.

Como mencionei no início desta seção, a palestra vai da célula até *healthcare*. Portanto, na sequência é abordado o mecanismo genético, de como a computação está afetando nessa parte etc. Obviamente não vou descrever cada um dos problemas da área. São aproximadamente 200 slides de conteúdo e caso você tenha

interesse por um problema em específico, com dezenas de aplicações e algoritmos sugeridos, basta [consultá-los](#).

Vou avançar para parte de medicina na qual tenho mais interesse. Assim como na biomedicina, elas apontaram diversas oportunidades na área. Destaco algumas:

- Diagnóstico e avaliação (aprendizado supervisionado)
- Auxiliar hospitais e clínicas a funcionar de maneira mais eficiente (pesquisa operacional)
- Intervenção de paciente em tempo real (otimização)
- Ajudar evitar erros humanos (ética, leis e bias)
- ML4Health em clínicas ruins e em casa (adaptação de domínio)
- Entender imagens e linguagem (visão computacional e NLP)

O ciclo completo de para entender e prever uma doença baseado em dados clínicos:

- **Triagem:** o paciente possui um alto risco de possuir a doença?
- **Predição:** o paciente vai possuir essa doença num futuro próximo?
- **Diagnóstico:** o paciente possui essa doença agora?
- **Monitoramento:** a doença está avançando?
- **Tratamento:** qual dosagem de qual droga esse paciente deve receber?

Primeiro: **modela o estado do paciente**; Segundo: **análise o caso**.

Vários exemplos são dados, como: predição de doença no pulmão, de doença cardíaca, sepsis, retinopatia diabética, dermatologia, etc. Além disso, elas apresentam lacunas no entendimento desses problemas:

- Dados faltantes
- Falta de contexto
- Nem todas as doenças podem ser prevenidas uma vez que predita
- Como dizer para um paciente que ele tem risco de ter uma doença que foi predita via ML?
- **Demografia do paciente é muito importante!**
- Não existe *ground truth labels* em alguns problemas (Ex: radiologia)

Para concluir, elas apresentam alguns desafios de ML na medicina. E começam com um exemplo:

1. **Realidade:** pacientes com asma possuem pneumonia e é tratado de maneira mais agressiva. Poucos pacientes com asma morrem de pneumonia
2. **O que o modelo aprendeu:** se você possui pneumonia, é melhor de você tiver asma também.

De fato, isso está nos dados, mas não faz sentido clinicamente.

Resumindo os principais problemas:

- Falta de grandes conjuntos de validação. É difícil obter dado, logo também é difícil validar.
- Bias e baixa representatividade de certos grupos populacionais. Isso pode gerar algoritmos racistas.
- Sensibilidade a ruído. Com pouco ruído a classificação pode mudar.
- Artefatos nas imagens também geram bias e devemos tomar bastante cuidado.
- Taxas altas de falso positivo.
- Falta de interpretabilidade/explicabilidade dos modelos.

Desse grupo, elas destacaram principalmente bias, na qual é descrito como um dos maiores problemas. Muitas das vezes grupos populacionais são tratados de maneira diferente e a gente não possui dado suficiente para isso. Elas dão o seguinte exemplo: um paciente chega para ser tratado com dor no peito, nos ombros e suando. Deve ser checado a condição de ataque cardíaco. Se esse paciente for uma mulher, os dados são diferentes, pois os sintomas são diferentes e a faixa etária também. Mas é um grande desafio descobrir isso pelos dados. Muitas vezes um sistema é treinado com um conjunto de dados que representa uma população e colocado em execução em outro tipo de população. Além disso, muitos sintomas são raros e o modelo pode não conhecer. **Essa é um desafio aberto da área.**

3.3 How to know - Celeste Kidd

Essa foi a palestra da cerimônia de abertura da conferência e foi ministrada por uma psicóloga e em geral foi uma apresentação de como o cérebro utiliza o núcleo cognitivo para guiar o seu aprendizado no mundo. Existe tanto conhecimento no mundo que ninguém é capaz de absorver. Isso significa que indivíduos precisam escolher e cada escolha é muito importante porque vem acompanhada de um custo. E isso ocorre desde os primeiros meses de vida. E segundo ela, os dados que um bebê utiliza para aprender não se parece com os tipos de dados que utilizamos em ML. Nada é rotulado e é bem diferente da ImageNet, por exemplo. Ela apresenta um software que ela utiliza no laboratório dela para detectar para onde um bebê está olhando e colocando a atenção dele. No caso, escolhendo o que aprender. Entender como bebês e crianças fazem isso, é essencial para desenvolver uma melhor AI.

O mais importante dessa palestra foi que ela apresentou **5 coisas que todos que trabalham com ML deveria saber sobre seres humanos.**

1. **Seres humanos formam crenças continuamente.** Existe uma atualização contínua de conhecimento. Mesmo para coisas que achamos que não, mas estamos continuamente atualizando nossa crença sobre ela. Ex: sabemos o que é uma cadeira mas a todo momento que encontramos uma nova instância, estamos atualizando nosso conhecimento sobre.
2. **Convicção diminui o interesse.** Em outras palavras, não somos curiosos em relação a coisa que achamos que sabemos. Temos processamento limitado,

e essa é uma forma otimizada de absorver mais. Mais do que isso, se você acha que sabe a resposta para uma determinada pergunta, você não tem propensão a descobrir se você está certo. Além disso, se você estiver errado e alguém lhe corrigir, você tem propensão a esquecer esse feedback.

3. **Convicção é guiada por feedback.** Imagine um experimento que apresenta para você um conceito novo e você tem que responder se uma determinada instância pertence a esse conceito. Se você recebe um feedback positivo a cada resposta, não importa se você está certo ou não, você ganha mais confiança e convicção sobre o que você está aprendendo. Se você começa a imaginar que a terra é plana e obtém alguma "evidência" no YouTube isso te faz criar convicção sobre esse novo conceito que você aprendeu. Muito embora essa seja uma ideia absurda, o fato de você receber um feedback positivo, te faz ter uma super convicção sobre esse assunto. Obviamente isso é problemático e estamos presenciando isso.
4. **Pouco feedback pode encorajar excesso de confiança.** Um experimento com pessoas com duas palavras diferentes no mesmo contexto que na verdade significa a mesma coisa mostram que pessoas confiam mais na sua resposta mesmo que a outra significa a mesma coisa.
5. **Seres humanos criam crenças rapidamente.** Ao ser exposta a uma informação, sendo verdade ou não, a pessoa vai criar uma crença sobre o assunto rapidamente. Atualmente, com a quantidade de informação que pode ser coletada online, as pessoas estão criando crença e não importa a qualidade dessa informação. Um experimento mostra que se uma pessoa busca a palavra "*activated charcoal wellness*", assistindo até 3 min de vídeo a em torno de 50% das pessoas já acreditam que isso é bom pra saúde dela, não importa se existe ou não evidência científica. Entre 7 a 10, mais de 90% acreditam. Obviamente, esse experimento foi realizado com pessoas sem background no assunto. Mas é isso que ocorre no dia a dia com os mais diversos assuntos, desde vacina até terra plana.

Compilando todas essas informações, a mensagem mais importante é: **não existe plataforma neutra.** Sempre que você produz uma tecnologia que entrega informação para as pessoas os algoritmos que rodam essa plataforma vai influenciar as pessoas para algum lado. E mais de 100 anos de pesquisa em psicologia mostram que a **maneira que a informação é apresentada determina a crença que a pessoa vai ter.**

Algoritmos estão moldando as nossas crenças e o nosso comportamento. Todos nós criamos conceitos online, mas os mais vulneráveis são as crianças. Elas criam crenças ainda mais rápido do que adultos. E isso é razão para uma preocupação séria do que está ocorrendo atualmente. **Todos os algo-**

ritmos que controlam essas redes tem profundo impacto nas crenças das pessoas.

Para acessar a palestra completa, [clique aqui](#).

3.4 Mais fontes interessantes

Como disse, apresentei minhas principais anotações do dia. Outras que eu recomendo ou que desejo ainda assistir e não consegui por conta dos eventos serem paralelos, estão listadas a seguir:

- Painel de discussão sobre estudantes latinos em ML. Este eu assisti e recomendo. Os professores falam sobre como obter sucesso na área, as desvantagens que nós temos e como devemos proceder em relação a isso. [Link para o painel](#).
- Tutorial: *interpretable comparison of distributions and models*. Eu gostaria muito de ter assistido esse mas foi ao mesmo tempo do tutorial de ML para biologia de medicina. Só consegui ver os slides superficialmente e pretendo assistir esse tutorial com cuidado nos próximos dias. [Link para o tutorial](#)
- Artigo: *Algorithmic Injustices: Towards a Relational Ethics*. Esse artigo ganhou o prêmio de melhor artigo do *workshop* Black in AI e discute a questão de bias e injustiça que algoritmos estão criando atualmente. [Link para o artigo](#).

4 Dia 3: Apresentações orais e pôsteres I

Neste dia começou a parte principal da conferência com as apresentações orais e os pôsteres. A maioria dos artigos (1228 de 1428) foram aceitos como pôsteres, que são apresentados em duas sessões de 2h, pela manhã e pela tarde (os pôsteres da manhã são diferentes dos da tarde). As apresentações orais são divididas em 4 *tracks* paralelas. Cada apresentação oral tem 5 min, a não ser que ela seja uma *spotlight*, que neste caso tem 15. Além disso, existem duas palestras principais. Uma pela manhã e outra pela tarde.

Compartilho aqui duas críticas em relação a esse modelo. Primeiro, 5 min é muito pouco tempo para apresentar qualquer coisa. O que da pra pegar é basicamente a ideia do artigo. Segundo, a NeurIPS valoriza muito mais o pôster do que outras conferências. Porém, uma sessão de 2h foi muito pouco tempo para expô-los. Por conta do alto número de pessoas para deslocar de uma palestra/sessão oral para área dos pôsteres era algo em torno de 15 a 20 min. Além disso, o local em que os pôsteres estavam expostos ficava extremamente lotado. Era muito difícil conseguir conversar com autor. Foram pouquíssimas vezes que eu consegui. Acredito que poderia ser mais tempo que aí sim era possível um contato maior.

4.1 Apresentações orais

Deep equilibrium models

Autores: Shaojie Bai, J. Zico Kolter e Vladlen Koltun. [Link para o artigo](#).

Neste artigo os autores apresentaram provas que eles podem trocar varias camadas de uma rede neural profunda por uma camada única sem perder capacidade representacional. Para isso eles propõe um Deep Equilibrium (DEQ) model. Segundo eles, esse modelo funciona igual ou melhor a modelos do estado da arte com custo de memória constante.

Eles apresentam um conceito chamado *weight-tied inputted-injected networks* (WTIIN), ilustrado na Figura 6. Em geral, como a mostrado na figura, parâmetros são compartilhados entre as camadas da rede. Eles demonstram teóricamente e empiricamente que a redução proposta não afeta no desempenho. Segundo eles, qualquer modelo de DL pode ser representado como uma WTIIN.

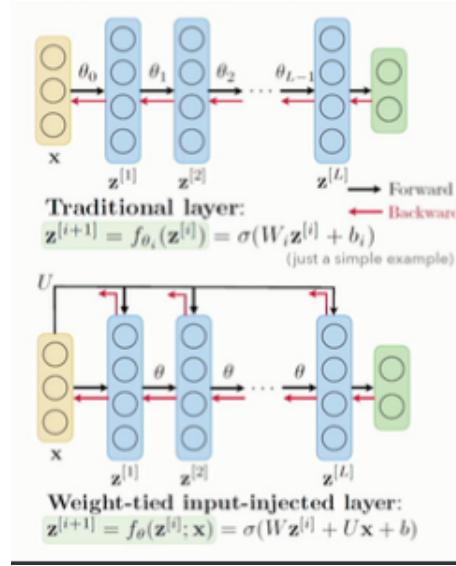


Figura 6. Ilustração do conceito *weight-tied inputted-injected networks* ([fonte](#))

Utilizando DEQ uma rede profunda pode ser imaginada como repetidas aplicações de alguma função:

$$\mathbf{z}^{[i+1]} = f_{\theta}(\mathbf{z}^{[i]}; \mathbf{x}) \quad (1)$$

A ideia é que o modelo converta para um equilíbrio:

$$\mathbf{z}^* = f_{\theta}(\mathbf{z}^*; \mathbf{x}) \quad (2)$$

A DEQ encontra esse equilíbrio via root-finding (quasi-Newton methods) ao invés de iterar no modelo. Um resumo dessa operação é exemplificado a seguir. Dado uma camada única $f_{\theta}(\mathbf{z}; \mathbf{x})$

1. **Forward pass:** dado uma entrada \mathbf{x} encontrar o ponto de equilíbrio \mathbf{z}^* tal que:

$$f_{\theta}(\mathbf{z}^*; \mathbf{x}) - \mathbf{z}^* = 0 \quad (3)$$

utilizando qualquer *root solver*.

2. **Backward pass:** implicitamente diferenciar através do estado equilibrado para formar os gradientes:

$$\frac{\partial l}{\partial(\cdot)} = \frac{\partial l}{\partial \mathbf{z}^*} \left(I - \frac{\partial f_{\theta}}{\partial \mathbf{z}^*} \right)^{-1} \frac{\partial f_{\theta}}{\partial(\cdot)} \quad (4)$$

$\frac{\partial f_{\theta}}{\partial(\cdot)}$ é o gradiente de uma camada e $\left(I - \frac{\partial f_{\theta}}{\partial \mathbf{z}^*} \right)^{-1}$ é a matriz Jacobiana no equilíbrio.

Essa operação vai deixar o treinamento e a inferência até 2 vezes mais lenta, porém, reduz o consumo de memória em até 90%. Para assistir a apresentação [clique aqui](#).

Towards explaining the regularization effect of initial large learning rate in training neural networks

Autores: Yuanzhi Li, Colin Wei e Tengyu Ma. [Link para o artigo](#).

Os autores observaram que utilizar uma *learning rate* (LR) inicial é importante para generalização. Neste artigo eles tentam explicar o porque.

Um *schedule* comum para treinar um modelo é começar com um valor alto para LR e utilizar uma queda constante (*annealing*). Porém, utilizar uma LR pequena é melhor do que uma grande até o início a operação de *annealing*, como ilustrado na Figura 7. A LR grande performa melhor apenas após a operação.

A explicação que eles tem para essa característica:

- LR pequena memoriza rapidamente padrões difíceis de ajustar. Ela ignora padrões, o que prejudica a generalização
- LR grande + *annealing* aprende padrões fáceis primeiro. Só memoriza padrões difíceis após o início do *annealing*. Isso ajuda na generalização.

Eles modificaram a CIFAR10 e para executar alguns experimentos. A conclusão deles foi que: LR pequena memoriza pequenas partes da imagem e ignora o resto. Por outro lado, uma LR grande ignora essas pequenas partes, que só serão aprendidas após o *annealing*. Por isso ela ajuda na generalização. Para assistir a apresentação [clique aqui](#).

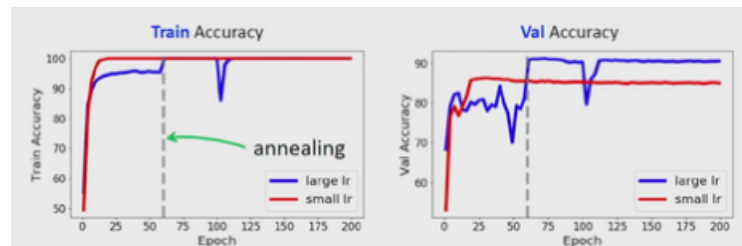


Figura 7. Desempenho dos modelos de acordo com o tamanho da LR

4.2 Pôsteres

Como já disse, a sessão de pôsteres estava extremamente lotada e era muito difícil conversar com os autores. Aqui vou citar alguns pôsteres que eu vi, gostei e vale a pena ler o artigo:

- **A Benchmark for Interpretability Methods in Deep Neural Networks.** Propõe uma métrica para medir a importância de uma feature em um modelo profundo. [Link para o artigo.](#)
- **Deliberative Explanations: visualizing network insecurities.** Uma nova técnica para visualizar o que um modelo profundo está aprendendo. [Link para o artigo.](#)
- **Information Competing Process for Learning Diversified Representations.** Apresenta uma nova forma para diversificar e aumentar o aprendizado de um modelo profundo. [Link para o artigo.](#)
- **q-means: A quantum algorithm for unsupervised machine learning.** Apresenta um novo algoritmo para clusterizar baseado em computação quântica. A ideia é aumentar a velocidade de computação do algoritmo. [Link para o artigo.](#)
- **Adversarial Examples Are Not Bugs, They Are Features.** Apresenta uma teoria de que o problema de ataques adversariais ocorrem por conta de features não robustas que não são compreendidas por um ser humano. [Link para o artigo.](#)
- **HYPE: A Benchmark for Human eYe Perceptual Evaluation of Generative Models.** Este artigo propõe uma maneira de medir qual modelo generativo é melhor do que o outro. [Link para o artigo.](#)

5 Dia 4: Apresentações orais, pôsteres II + evento satélite UBC

Neste dia, além da mesma sequência do dia anterior, ocorreu um evento satélite chamado [A UBC NeurIPS meetup: Biomedical imaging & AI](#) foi oferecido pelo [Biomedical Imaging and Artificial Intelligence Research Cluster \(BMIAI\)](#) da *University of British Columbia* (UBC). Como é a área que estou trabalhando atualmente, me inscrevi e fui participar deste evento, que contava com 3 palestras e um meetup ao final.

Também neste dia, ocorreu a palestra do Yoshua Bengio na NeurIPS. Infelizmente eu não pude acompanhar in loco pois era ao mesmo tempo do evento da UBC. Porém, mais tarde eu assisti a ela no sistema de streaming da NeurIPS.

5.1 Machine Learning in Medical Imaging : Progress So Far and Perspectives - Thad Hughes

Thad Hughes é um engenheiro de software da Google e nessa palestra ele apresentou os avanços da empresa na área médica. A maior parte da palestra foi de um trabalho recentemente publicado na Nature: [End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography](#). Infelizmente essa palestra (e nem uma da UBC) foi gravada e a apresentação não está disponível.

Para começar, as imagens utilizadas aqui são 3D, o que dificulta ainda mais o problema. O modelo que ele apresentou é dividido em duas partes. A primeira apenas detecta a ROI, ou seja, onde está o tumor na imagem. A segunda realiza a predição do risco de câncer. A Figura 8 mostra algumas ROIs que o sistema extrai.

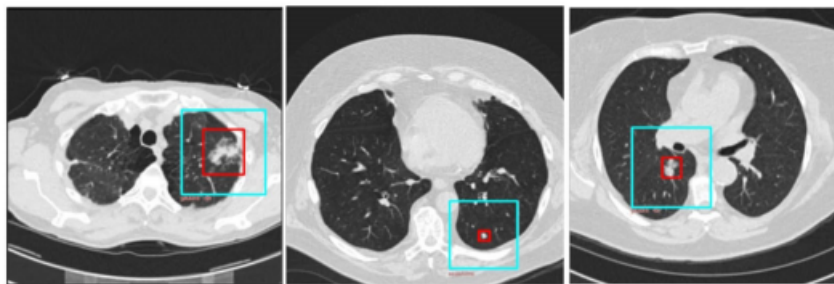


Figura 8. Exemplo de uma ROI extraída pelo sistema ([fonte](#))

Uma visão geral desse modelo é ilustrada na Figure 9. Nessa figura é interessante observar que eles agregam features. Essa agregação é realizada via concatenação.

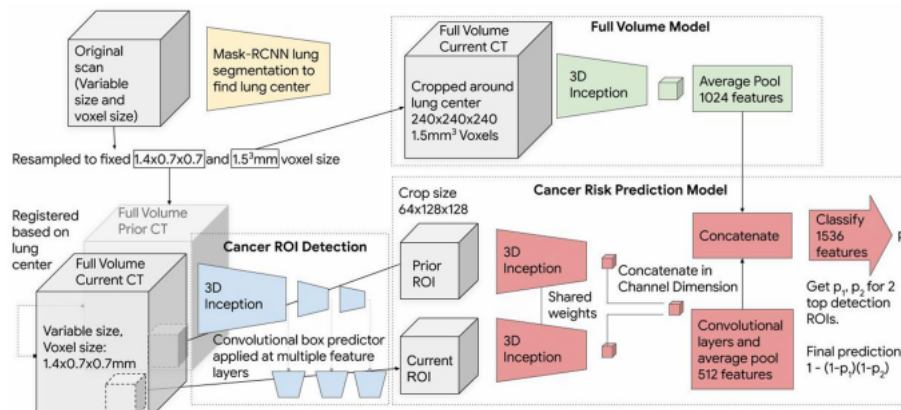


Figura 9. Arquitetura do modelo proposto para predição de câncer de pulmão (fonte)

Alguns dados em relação ao treinamento do modelo que foi apresentado: eles utilizaram 70% dos dados para treino, 15% para validação e 15% para teste. Porém, eles possuem um outro dataset para testar o modelo. Segundo ele isso é importante para variar o conjunto populacional que o dataset é gerado. Isso vai ao encontro do tutorial descrito na seção 3.2. Além disso, eles também tem problemas de dados desbalanceados e em uma pergunta ele respondeu que eles utilizaram *Focal Loss* para lidar com esse problema.

Como todo problema médico, eles tiveram problemas com bias. Mas um deles foi inusitado. Algumas imagens continha as iniciais do radiologista que solicitou. Obviamente, entre os radiologistas, havia desempenhos diferentes. O modelo identificou essas iniciais, aprendeu quais eram as iniciais que possuíam melhor desempenho e começou a usar essa informação para prever a lesão. Eles tiveram que tratar esse problema. Não ficou muito claro, mas pelo o que eu entendi eles removeram as iniciais das imagens.

Eles também desenvolveram uma ferramenta pra identificar as features que contribuem positivamente e negativamente para um câncer. O resultado é uma imagem como mostrado na Figura 10. Nessa imagem, em roxo são as features positivas e em azul as negativas. Houveram algumas perguntas em relação a geração dessa imagem, mas o palestrante não tinha muitas informações, uma vez que ele não participou especificamente nessa parte do projeto.

Para concluir, a performance deste modelo está superior a de radiologistas estabelecendo um novo marco para área. A AUC para predição de câncer em 1 ano está em 0.950 e para 2 anos em 0.882.

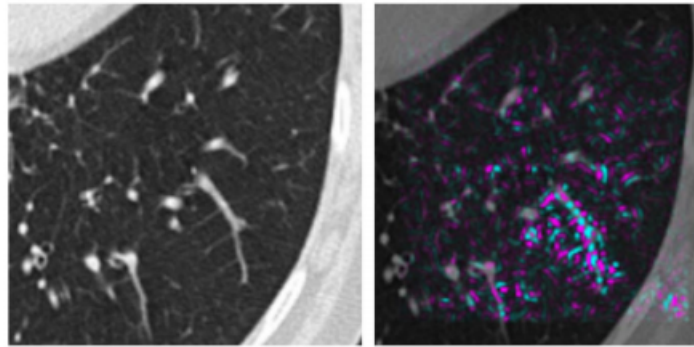


Figura 10. Interpretabilidade do modelo de predição de câncer de pulmão ([fonte](#))

5.2 Deep learning for intelligent medical imaging solutions - Julia Schnabel

Julia Schnabel é diretora do Centre for *Doctoral Training in Medical Imaging*, que é dirigido pela *King's College London* e a *Imperial College London*. Ela apresentou diversos resultados recentes do grupo dela, incluindo:

- [Weakly Supervised Estimation of Shadow Confidence Maps in Fetal Ultrasound Imaging](#). CVPR18
- [Quantitative assessment of myelination patterns in preterm neonates using T2-weighted MRI](#). Nature Scientific Reports.
- [Detection and Correction of Cardiac MR Motion Artefacts during Reconstruction from K-space](#). MICCAI19.

O trabalho dela envolve mais *Magnetic Resonance Imaging* (MRI), ou seja, como reconstruir dados do domínio da frequência para imagens. Na sequência, como analisar e trabalhar com essas imagens. Segundo ela, a experiência dela mostra que o sucesso de DL para esse tipo de imagem é limitado não só pela **quantidade**, mas também pela **qualidade** das imagens obtidas.

Ela compartilhou um problema semelhante com o caso das iniciais dos médicos nas imagens do câncer de pulmão. Neste caso, são artefatos médicos que saem nas imagens e podem causar bias. Eles utilizaram uma RCNN para identificar e depois remover esses artefatos. A arquitetura dessa rede é ilustrada na Figura 11.

Outros problemas que ela destacou, foi desbalanceamento de dados. Neste caso, eles utilizam *data augmentation* para lidar com o problema. Além disso, ela lista como desafios: desenvolver um modelo para detectar imagens de baixa qualidade e descartá-las do processo de treinamento e de decisão. E de melhorar modelos de interpretação e explicação do que a rede está aprendendo. Novamente, ao encontro do tutorial da seção 3.2.

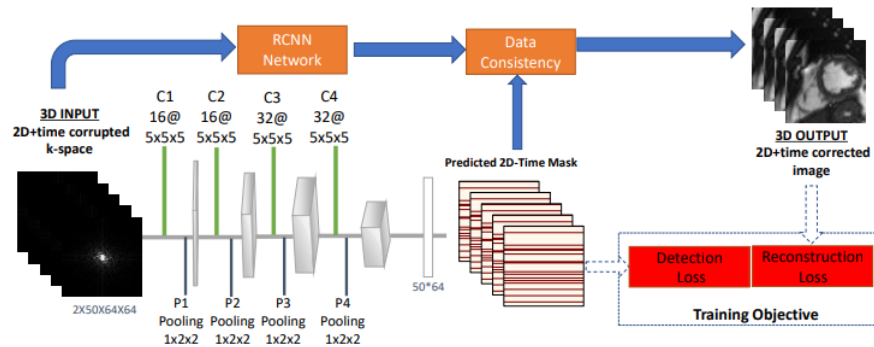


Figura 11. Arquitetura da RCNN utilizada para remover artefatos de MRI ([fonte](#))

5.3 AI in healthcare working towards positive clinical impact - Nenad Tomašev

Nenad Tomašev é um pesquisado associado ao *DeepMind*, Google. A palestra dele foi centrada basicamente no artigo publicado recentemente na Nature: [A clinically applicable approach to continuous prediction of future acute kidney injury](#).

O modelo que eles propuseram para atacar este problema é ilustrado na Figura 12. De maneira geral o modelo utiliza dados do histórico do paciente e da atual situação. São extraídas features via camadas totalmente conectadas que são enviadas para redes recorrentes. Na sequência ocorre uma agregação via concatenação e os resultados são coletados.

Dois me chamaram a atenção na apresentação:

- Eles possuem um dataset com 703,782 amostras de pacientes adultos. Neste caso, não são dados de imagens.
- Eles realizam uma busca por parâmetros que basicamente demanda uma quantidade enorme de infraestrutura. Eles variam basicamente todos os parâmetros do modelo para encontrar o que devolve o melhor desempenho.

Como de praxe, ele disse que os dados dele possui forte bias e representa apenas um pequeno espectro da população. Mais especificamente, homens brancos, uma vez que o dataset é coletado por meio do *US Department of Veterans Affairs*.

5.4 From System 1 Deep Learning to System 2 Deep Learning - Yoshua Bengio

A área de *deep learning* avançou significativamente no último século. Porém, uma pergunta importante deve ser feita: **será que é suficiente apenas aumentar o tamanho dos datasets, dos modelos e a velocidade computacional?** Um cérebro maior é suficiente? Na visão de Bengio, não.

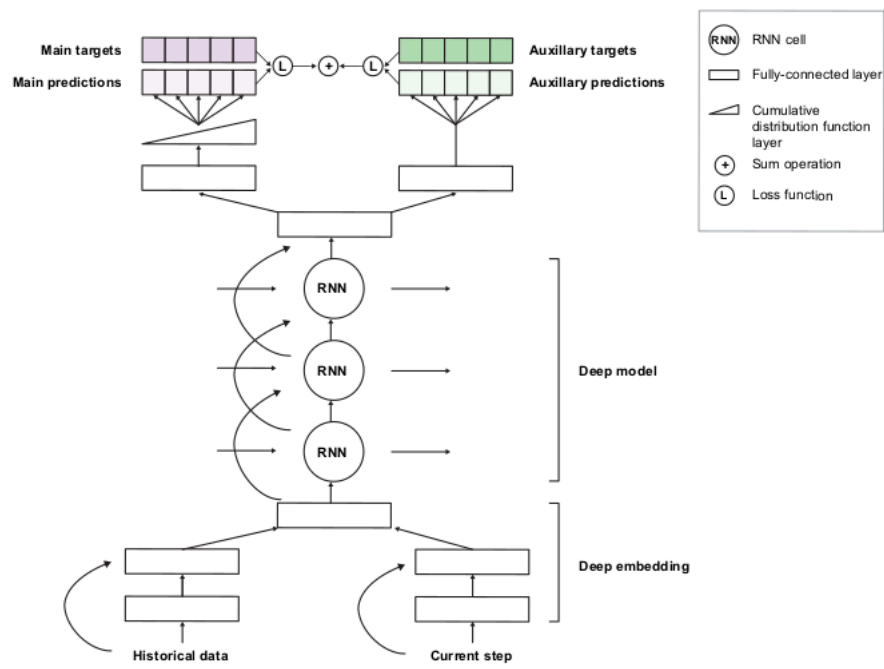


Figura 12. Arquitetura do modelo do modelo proposto ([fonte](#))

AI ainda está longe do nível humano. Ainda falta eficiência, os métodos são limitados, falta robustez, erros estúpidos ocorrem. Alguns pesquisadores acreditam que devemos pensar em algo completamente diferente de DL para atingir uma cognição de mais alto nível. Porém, nesta palestra Bengio apresenta um caminho para o que ele chama de *system 2*.

A inspiração para esse caminho veio do livro [Thinking fast and slow](#) de Daniel Kahneman (para quem estava procurando algo para ler, essa é a dica).

Mas o que é *system 1 and 2*:

- **System 1:** são tarefas intuitivas, rápidas, **inconsciente**, habitual e não linguísticas. São tarefas que atualmente DL desempenha bem. Ex: Se você está dirigindo um carro a caminho para sua casa, você não precisa prestar atenção no caminho. É intuitivo.
- **System 2:** são tarefas lógicas, lentas, sequenciais, **consciente**, linguística, planejada. Precisa manipular alto nível de conceitos linguísticos que precisam ou não ser combinados com visuais. Isso é o futuro de DL. Ex: Se você está dirigindo para um local desconhecido, você precisa prestar atenção em placas, esquinas, referências visuais, etc.

Bengio apresenta alguns conceitos que ele acredita que devemos melhorar com intuito de produzir AI no nível humano:

- Precisamos trabalhar melhor em *out-of-distribution generalization e transfer learning*
- Cognição em alto nível: é necessário evoluir do sistema 1 para o 2. Para isso é necessário trabalhar com modelos semânticos de alto nível que seja capaz de compor conceitos que nós humanos manipulamos com a linguagem. E muitos desses conceitos tem a ver com causalidade, ou seja, tem causa e efeito.
- E quando se fala de causalidade, é necessário trabalhar com *agent perspective*¹, algo que o *mainstream* de DL não tem colocado muita atenção.

Esses 3 tópicos estão conectados entre si. Entender como eles estão linkando é fundamental para traçar o caminho do sistema 1 para 2.

Os principais pontos que ele acredita que devem ser tomados para potencializar o sistema 2:

1. Objetivos em ML: lidar com mudanças na distribuição dos dados, que algo necessário para agentes, uma vez que não são estacionários.
2. Blocos básicos para o sistema 2: *attention & consciousness*. Conceitos chaves para alcançar o próximo nível.

¹ Este é um conceito relacionado a tomar uma ação e ter total controle sobre essa ação. Em outras palavras, *agents* está relacionado com ações. Algum modelo de aprendizado que está embarcado em algum ambiente.

3. *Consciousness prior*: a distribuição conjunta dos conceitos de alto nível pode ser representada por uma *sparse factor graph*.
4. *Framework teórico*: como o mundo muda que se conecta com a ideia de *agents*. Na maioria das vezes quando coisas mudam no mundo é por conta de *agent* (como pessoas) faz algo.
5. Composição de arquiteturas DL que o processamento do sistema 2 necessita. Neste caso, ele acredita que os modelos não deve apenas trabalhar com vetores mas sim conjuntos de objetos

Lidando com mudanças na distribuição A hipótese clássica da teoria em ML é baseada em *Independent Identically distribution* (IID) data, ou seja, **a distribuição de treino é a mesma da de teste**. Se nós não fizermos essa suposição nós não podemos dizer nada sobre generalização. Porém, essa suposição é muito **forte**. Na prática, isso não ocorre.

Uma prática comum que todos fazem é embaralhar os dados quando o coletamos. Porém:

Natureza não embaralha dados, então nós não devemos fazer isso
L. Bottou, ICML19

E a razão pela qual nós não devemos fazer isso é porque destruímos informações importantes sobre as mudanças na distribuição que são herdadas dos dados que nós coletamos. E ao invés de destruímos essas informações, nós deveríamos utilizá-las para entender como o mundo muda.

O problema de OOD ocorre por conta da nossa suposição de IID. Na verdade, OOD quebra essa suposição. Sendo assim, é necessário trocar essa suposição por alguma outra coisa para sermos capazes de lidar com OOD.

Agent learning necessita de generalização de OOD². Isso é necessário porque *agents* quase sempre precisa lidar com dados não estacionários. Existem mudanças na distribuição por diversos motivos: ação de outros agentes, por conta de lugares diferentes, sensores, regras, etc.

Na visão dele, um conceito chave para atingir a generalização de OOD é introduzir mais formas de *compositionality*. Mas o que isso significa? Em resumo, significa ser capaz de dinamicamente recombinar conceitos existentes, **mesmo quando esses conceitos não tem chance de acontecer**. Ele dar um exemplo de filmes de ficção científica. Nós não conhecemos e não vivemos em um ambiente como aquele. Mas nós podemos imaginar e ter uma ideia significativa de como é. Ele chama isso de **systematic generalization**. **Os modelos atuais não são capazes de lidar com isso.**

² Basicamente a necessidade do modelo generalizar algo de uma distribuição diferente

Ok, mais como isso difere da AI clássica? Segundo ele, isso evita problemas clássicos em AI relacionados a manipulação de regras baseadas em símbolos. Porém, é necessário:

- Aprendizado eficiente em larga escala
- Atrelar semântica no sistema 1
- Representações distribuídas para generalização
- Mais eficiência no treinamento do sistema 1
- Lidar com incerteza do mundo real
- *Systematic generalization*
- Fatorizar o aprendizado em partes menores
- Manipular variáveis

Blocos básicos para o sistema 2: *attention & consciousness* Para começar, o que é *attention*?

Realizar computação de uma maneira focada. Sequencialmente focar a computação em um elemento de cada vez. Isso é extremamente poderoso. Por exemplo, quando estamos fazendo uma tradução de uma língua para outra, estamos focando em uma sentença mais curta para fazer a tradução de forma mais eficiente.

Ele propões um método chamado **Content-based soft attention**: nós podemos aprender onde ter atenção. Além disso, é útil pois é possível *backpropagate* para este local. Em resumo, imagine várias camadas de uma rede. A camada subsequente decide onde colocar atenção da camada anterior baseado em um *score* de onde ela deve colocar atenção. "*Attention* é uma ação interna. É necessário uma *learned attention policy*". *Attention* vem sendo usada em diferentes áreas:

- NLP: *self-attention, transformers*
- Atacando o problema de *vanishing gradients*
- *Memory-extended neural nets*

Attention cria uma conexão dinâmica entre duas camadas de uma rede neural. A ideia é que a informação que vai propagar para a próxima camada, deve ser escolhida de acordo com a *attention*. Ainda é um desafio propagar essa informação.

From attention to consciousness: existe uma teoria chamada *Global workspace theory*:

- No cérebro existe um seletor de informação
- Alguns elementos que são processados no nosso cérebro são selecionados e espalhados para o resto do mesmo de tal forma que influência as outras partes.
- Isso é relacionado com uma memória curta na qual esses elementos são selecionados e colocados a disposição para as demais partes.

Consciência está relacionada com língua. Uma maneira que sabemos que alguém tem consciência é perguntando o que ela pensa sobre determinado assunto. Existe um forte link entre nossos pensamentos e nossa consciência.

Consciousness prior: sparse factor graph Ele propões um modelo em um artigo chamado [The Consciousness Prior](#). Em resumo, é um modelo de *sparse factor graph* em que os nós são variáveis e as bordas representam relações entre as variáveis. Segundo ele, isso faz sentido pois é similar a maneira que utilizamos nossa língua. Nós conseguimos prever próximas palavra dada as anteriores. Porém, isso não funciona para imagens. Não é possível encontrar um conjunto pequeno de 5 pixels uma imagem de modo que possamos prever com acurácia alta um pixel dados os outros 4. Mais isso é útil para NLP.

Framework teórico Ele começa essa parte citando Charles Darwin:

"Não é a espécie mais forte ou mais inteligente que sobrevive, mas sim a mais reponsiva a mudanças."

Segundo ele o ser humano está em contínuo aprendizado. Algo na linha da palestra *How to know*. Logo, técnicas de *meta-learning*, ou seja, aprender a aprender, são necessárias. Ele pensa em um modelo com dois níveis de otimização. Um interno que retorna uma *loss* e outro externo de aprendizado contínuo que otimiza essa *loss*. Lembra um pouco a *gameificação* de uma GAN.

Composição de arquiteturas em DL Ele cita um modelo chamado [Recurrent Independent Mechanisms \(RIMs\)](#) que é capaz de operar em conjuntos e objetos além de modularizar os cálculos computacionais. O modelo possui múltiplos módulos esparsos recorrentes que resulta em uma generalização melhor de OOD.

5.5 Pôsteres

- **Control Batch Size and Learning Rate to Generalize Well: Theoretical and Empirical Evidence.** Este paper mostra que a capacidade de generalização está negativamente correlacionada com tamanho do *batch* e positivamente com a taxa de aprendizado. [Link do artigo](#).
- **Small ReLU networks are powerful memorizers: a tight analysis of memorization capacity.** Este artigo traz um teorema que diz que qualquer FNN com 3 camadas ReLU, no caso 2 *hidden layers* com dimensão d_1 e d_2 , pode ajustar qualquer dataset se $d_1 d_2 \geq 4N d_y$, sendo N o número de amostras e d_y a dimensão da saída. Para entradas escalares, $d_1 = d_2 = 2\sqrt{N}$. [Link do artigo](#).
- **Self-Critical Reasoning for Robust Visual Question Answering.** Este artigo trabalha com parte do problema apresentado na palestra do Carlos Guestrin, descrita na seção 3.1. [Link do artigo](#).

- **Image Captioning: Transforming Objects into Word.** Este artigo propõe uma nova metodologia baseado em inter-relação espacial para rotular uma imagem. [Link para o artigo](#)

5.6 Mais fontes interessantes

A palestra [Social Intelligence](#) de Blaise Aguera y Arcas parece ser muito interessante. Infelizmente, não consegui assisti-la por completa in loco e pretendo fazer isso em breve. Porém, uma parte que me chamou atenção foi quando ele diz que a área de ML está trabalhando de maneira errônea a parte da otimização. O cérebro não avalia apenas uma função. O cérebro está se modificando o tempo inteiro e é capaz de aprender por experiência. A área precisa a pensar algo mais biológico que vá além de apenas utilizar o backpropagation para uma função.

6 Dia 5: Apresentações orais, pôsteres III

Mesma programação dos dias 2 e 3.

6.1 Agency + Automation: Designing Artificial Intelligence into Interactive Systems - Jeffrey Heer

Ele começa a palestra na linha de outras já resumidas aqui: a velha retórica de que AI vai substituir o ser humano. Ele é contra isso e acha que AI deve ser utilizada para colaborar. Ele cita Prof. Fei-Fei Li:

(...) Se desejamos afetar positivamente o mundo no futuro, AI deve ser guiada por interesses humanos... nos melhorando e não nos substituindo

Há 20 anos, Eric Horvitz discutia conceitos como este no seu artigo [Principles of Mixed-Initiative User Interfaces](#). Ele forneceu:

- Ele forneceu valores importantes a serem incluídos na automação
- Apoio incerteza em relação ao objetivo do usuário
- Mecanismos para colaboração para melhorar resultados
- Aprendizado contínuo para aprender ações do usuário

Existe uma pergunta fundamental: **qual é o equilíbrio entre automação e controle do usuário?** De todas as anotações anteriores, já sabemos de bias, erros estúpidos, falta de representatividade, etc. A proposta dele para essa questão é melhorar a interface do usuário utilizando as capacidades dos modelos, ações e objetivos para possibilitar uma maior interação entre o ser humano e AI. Exemplos:

- Exploração e visualização de dados
- *Data cleaning e Transformation*
- *Natural Language translation*

Exploração e visualização de dados A ideia aqui é ter insights em alto nível em relação aos dados de modo que auxilie o usuário a interpretar e entender melhor os dados que ele coletou. Por exemplo, clusterizar os dados e exibir em um gráfico.

Mas o que faz uma visualização boa? Obviamente essa é uma pergunta subjetiva e depende de quem estar respondendo. Porém existem conceitos e percepções que permite atacar essa questão. Para isso, ele faz um pequeno experimento mostrando um círculo grande e um pequeno e pergunta a audiência quantas vezes o círculo grande é maior do que o pequeno. Na sequência, ele apresenta duas barras, uma grande e uma pequena e faz a mesma pergunta. O teste mostra que a pessoas concordam mais com a barra do que o círculo³.

Em termos psicológicos, pessoas são boas em interpretar comprimento, altura, posição, inclinação... Porém, são ruins com cores, volumes e áreas, como ilustrado na Figura 13.

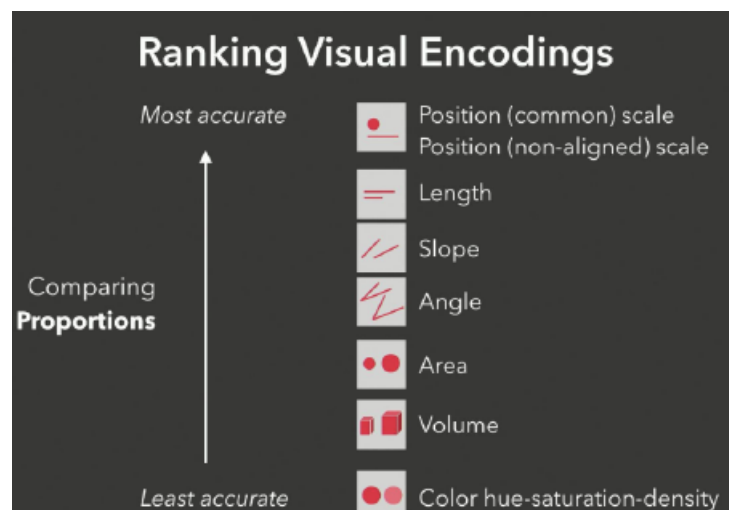


Figura 13. Ranking de visualizações mais acuradas ([fonte](#))

Ele apresentou uma ferramenta de exploração de dados chamada [Voyager](#), que é de código aberto e aparenta ser uma ferramenta interessante para a análise de dados. Mas o mais interessante é que você pode carregar os dados e a ferramenta já faz sugestões de visualizações, o que vai na linha de interação entre AI e

³ Esse experimento é bem legal e sugiro que assista. Começa por volta do [minuto 13 do vídeo](#).

ser humano. Segundo os resultados que obtiveram dos usuários, a ferramenta acelerou em mais de 2x a avaliação dos dados.

Data cleaning e Transformation Analistas de dados passam a maior parte do tempo preparando e limpando dados. Essa é uma tarefa extremamente dispendiosa.

Em Data Science, 80% do tempo é gasto preparando os dados e 20% reclamando que é necessário preparar os dados

Ele apresenta outra ferramenta do grupo dele chamada [Data Wrangler](#). A ideia é a mesma, interação entre AI e ser humano. O sistema vai aprendendo a limpar os dados e otimizar o tempo do analista. Ele prever algumas ações faz a sugestão e usuário deve tomar a decisão.

Novamente, os experimentos deles mostraram que o tempo gasto com essa tarefa foi reduzido além de apresentar boas sugestões que o analista poderia não perceber.

Natural Language translation Nest último tópico ele diz que esse problema era muito difícil de se ter colaboração porque a AI era muito ruim no passado. Porém agora as coisas evoluíram e ele cita outro trabalho do grupo dele chamado [predictive translation memory \(PTM\)](#)

Similar aos tópicos anteriores, eles aplicam isso para trabalhar colaborativamente com tradutores. Os resultados aqui são interessantes. Utilizar a ferramenta para traduzir completamente o texto e na sequência um ser humano editar essa tradução obteve ganho de eficiência e qualidade na tradução. Porém, para realizar essa tarefa interativamente, a tradução melhora mas o tempo aumenta. Usuários reportaram que ele se sente menos criativo ao utilizar a ferramenta pois ela o distrai e coloca palavras diferentes das que ele estava pensando. O desafio aqui é entender melhor como encaixar essa colaboração.

Desafios futuros

- **Desenvolver processos, ferramentas e monitoramento.** Como nós podemos apoiar o desenvolvimento, *deployment* e monitoramento de sistemas interativos que utilizam ML?
- **Mapper a representação aprendida pela máquina.** ML pode sugerir tarefas estruturais? Pessoas podem auxiliar com teste e limitações de modelos de ML?
- **Avaliar o *trade-off* entre essa interação.** É preciso ir além de resultados qualitativos e de produtividade.

Para ter acesso a palestra completa [clique aqui](#).

6.2 Apresentações orais

Adversarial Music: Real World Audio Adversary Against Wake-word Detection System

Autores: Juncheng B. Li, Shuhui Qu, Xinjian Li, Joseph Szurley, J. Zico Kolter, Florian Metze. [Link para o artigo](#).

Ataques adversariais é um problema em visão computacional. Porém, trabalhos recentes também mostram que este é um problema na área de áudio. O objetivo deste artigo é fazer uma ataque a assistentes de áudio como Alexa e Siri. Esses assistentes funcionam com uma *wake-word*, ou seja, você fala uma palavra, normalmente o nome da assistente, e ela começa a responder os seus comandos.

O primeiro problema é que esses assistentes são *black-boxes*. Mas existem algumas publicações dos desenvolvedores que os guiaram para desenvolver um "*Grey Box*" *attack*, ou seja, eles implementaram um versão deles do assistente para desenvolver o sistema de ataque.

Sendo assim, eles propuseram um sintetizador de som para gerar um ruído. Quando esse ruído está presente a acurácia de detecção da *wake-word* da Alexa cai de 93% para 11%, ou seja, o ataque foi bem sucedido. Eles também testaram diferentes dbA, distâncias e ângulos e o resultado persiste.

Eles disponibilizam uma demonstração [neste vídeo](#).

This Looks Like That: Deep Learning for Interpretable Image Recognition

Autores: Chaofan Chen, Oscar Li, Chaofan Tao, Alina Jade Barnett, Jonathan Su, Cynthia Rudin. [Link para o artigo](#).

Ao observar um pássaro, quais são os atributos que são utilizados para classificá-lo? Esse artigo propõe um modelo para olhar para *features* específicas que possam explicar a classificação desse pássaro, como ilustrado na Figura 14.

Eles propuseram um modelo chamado ProtoPNet, que se inspira na maneira que o ser humano identifica explicações para classificar algo. O modelo apresenta partes para interpretação mas também possui a explicação do porque daquela parte.

Os experimentos apresentam desempenho competitivo com SOTA.

6.3 Pôsteres

- **Paradoxes in Fair Machine Learning.** Este artigo apresenta resultados empíricos em relação equalização o problema de discriminação em relação alguns grupos sociais em problemas de classificação. [Link do artigo](#).

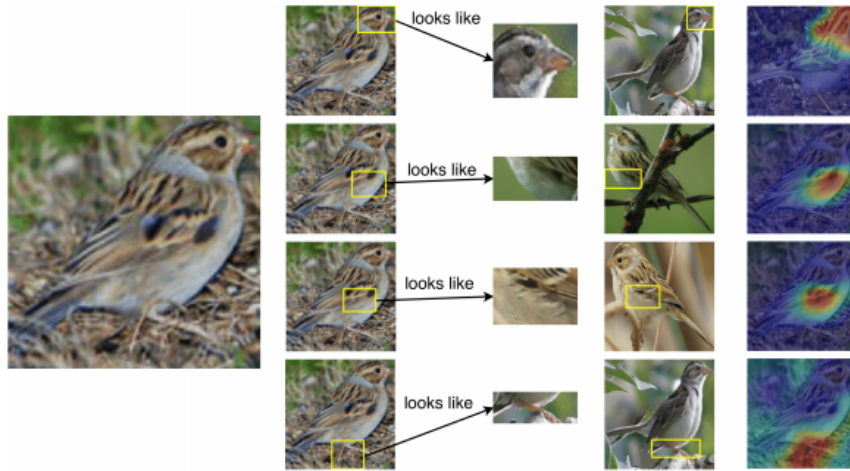


Figura 14. Exemplo de como o modelo identifica as features visualmente ([fonte](#))

- **Learning Representations for Time Series Clustering.** Esse artigo apresenta uma proposta para utilizar seq2seq para clusterizar séries temporais e obter uma representação efetiva da mesma. [Link para o artigo.](#)
- **Ask not what AI can do, but what AI should do: Towards a framework of task delegability.** Aborda mais uma questão ética da AI e propõe um framework para mensurar o quanto as pessoas aceitam de algo automatizado no seu cotidiano. [Link para o artigo.](#)
- **Transfusion: Understanding Transfer Learning for Medical Imaging.** Apresenta provas empíricas de que transferir apenas as primeiras camadas e todas as camadas de uma CNN resultam em resultados similares em dois datasets médicos. [Link para o artigo.](#)
- **Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss.** Intruz uma nova função de perda chamada *label-distribution-aware margin (LDAM) loss* para substituir a *cross-entropy* e uma abordagem para *re-weighting* após a fase inicial para permitir que o modelo aprenda a representação inicial dos dados de maneira mais adequada. [Link para o artigo.](#)
- **Defending Against Neural Fake News.** Propõem um modelo para identificar fake news gerados por modelos NLP. Também apresenta uma discussão ética em relação a esses modelos. [Link para o artigo.](#)

- **Stacked Capsule Autoencoders.** Uma versão modificada da CapsNet que usa explicitamente (ao invés de implicitamente) as relações geométricas entre as partes que explicam um objeto. [Link para o artigo.](#)
- **U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging.** Uma versão modificada U-Net que aplicada para sequencia de dados, como por exemplo, dados de sensores. [Link para o artigo.](#)
- **On Testing for Biases in Peer Review.** Um artigo super interessante e necessário que investiga bias em processos de revisão de conferências como a própria NeurIPS. [Link para o artigo.](#)

6.4 Mais fontes interessantes

A palestra [Mapping emotions: discovering structure in mesoscale electrical brain recording](#) de Kafui Dzirasa mostra como o laboratório dele adquire informações cerebrais em testes com rato de laboratório e utiliza isso para prever depressão, por exemplo.

7 Dia 6 e 7: workshops

Os dois últimos dias da conferência foram destinados aos *workshops*. Como disse na introdução, *workshop* na NeurIPS é quase que uma sessão especial sobre um determinado assunto. São mais de 50 *workshops* ocorrendo simultaneamente. Caso queira ver a lista completa dos workshops, você encontra [neste link](#). O workshop que eu passei a maior parte do tempo foi o que eu tive o meu artigo aceito, no caso, o [Retrospectives](#). Além deste, eu assisti algumas palestras e vi alguns pôsteres nos seguintes *workshops*:

- [ML4H: Machine Learning for Health](#)
- [ML for the Developing World \(ML4D\)](#)
- [Fair ML for Health](#)
- [Deep Reinforcement Learning Workshop](#)
- [Tackling Climate Change with Machine Learning](#)
- [Medical Imaging Meets NeurIPS](#)
- [Machine Learning for Creativity and Design](#)
- [AI for social good](#)

Cada um desses workshops tem seus próprios sites e lista de artigos que foram aceitos. Alguns deles também ocorrem na ICLM, ICLR, CVPR etc. Na sequência vou deixar linkado as palestras/paineis que eu vi e achei relevantes e a lista de artigos/pôsteres.

Palestras

- **Computational + Systems vs climate change.** Essa foi a palestra de abertura do *Tackling Climate Change with Machine Learning* e o palestrante foi Jeff Dean, o head do Google Brain. Nela ele apresenta as medidas e produtos que o Google está tomando e desenvolvendo para atacar a mudança climática. [Link para a palestra.](#)
- **Changing the Paradigm of Pathology: AI and Computational Diagnostics.** Essa palestra foi dada por Leo Grady, CEO da Paige AI, e aborda as dificuldades e como a empresa está trabalhando com AI para auxiliar patologistas. Dois pontos interessantes é que ele mostra como os patologistas erram no diagnóstico do câncer e como a alta dimensão das imagens é um desafio para se trabalhar nessa área. Por exemplo, a CIFAR10 inteira é do mesmo tamanho que algumas dezenas desse tipo de imagem. <https://slideslive.com/38922092/medical-imaging-meets-neurips-3>.
- **AI in radiology.** Palestra ministrada por Daniel Sodickson na qual é discutindo os avanços de AI/ML em radiologia. Uma parte interessante é que ele diz que AI não vai substituir os radiologistas, mas os radiologistas que usam AI vão substituir os que não usam. [Link para a palestra.](#)
- **Fast MRI challenge.** Apresentação feita por Larn Zitnick, pesquisado do Facebook Research, apresentando o desafio que o Facebook fomentou para esse workshop. O problema aqui é que MRI precisa ser convertida do domínio da frequência para uma imagem humanamente interpretável. O problema é que são muitos dados para converter e esse processo demora muito. O desafio é tentar acelerar isso sem perder qualidade na imagem gerada. A grande questão é: os detalhes da imagem são muito importante e quando se comprime o vetor da frequência para acelerar a conversão, eles são perdidos. [Link da apresentação.](#)
- Painel de discussão com Yoshua Bengio, Andrew Ng, Carla Gomes, Lester Mackey e Jeff Dean, com o tema: *Climate Change: A Grand Challenge for ML*. É interessante ver o ponto de vista de cada deles em relação a mudança climática e como a área deve encarar o problema.
- **Playing Dota 2 with Large Scale Deep Reinforcement Learning.** Apresentação descrevendo como a OpenAI utilizou Deep Reinforcement Learning para criar uma AI para ganhar dos campeões mundiais de Dota 2. Um ponto interessante aqui é que o modelo deles treino por quase 1 ano e eles tinha que fazer atualizações em tempo de treinamento na função de perda do modelo. [Link para apresentação.](#)

- **Bulshit that I and others have said about Neural ODEs.** Essa apresentação foi concedida por Davif Duvenaud, professor da UofT e pesquisador do Vector Institute. Como o nome dela sugere, ele descreve erros e falácias que ele cometeu em relação ao tema. [Link para apresentação.](#)
- Painele com Yoshua Bengio, Joelle Pineau, Melanie Mitchell, Gael Varoquaux, and Jonathan Frankle, discutindo problemas na área de ML, limitações que não são faladas, problemas que ocorrem nas revisões dos artigos e saúde mental de alunos de pós-graduação e pesquisadores recém formados. **Eu recomendo fortemente esse painele.** Toda conferência deveria ter algo do tipo. [Link para o painele.](#)

7.1 Pôsteres

Aqui vou listar os pôsteres que eu vi e que achei o tema interessante:

- [Creative GANs for generating poems, lyrics, and metaphors](#)
- [Immersion - How Does Music Sound to Artificial Ears?](#)
- [Immersion - How Does Music Sound to Artificial Ears?](#)
- [Localization with Limited Annotation for Chest X-rays](#)
- [Improved Hierarchical Patient Classification with Language Model Pre-training over Clinical Notes](#)
- [Federated and Differentially Private Learning for Electronic Health Records](#)
- <https://ml4health.github.io/2019/pages/extended-abstracts.html>
- <https://ml4health.github.io/2019/pages/extended-abstracts.html>
- [Detecting cutaneous basal cell carcinomas in ultra-high resolution and weakly labelled histopathological images](#)
- [Semi-Supervised Histology Classification using Deep Multiple Instance Learning and Contrastive Predictive Coding](#)
- [Deep Learning for the Digital Pathologic Diagnosis of Cholangiocarcinoma and Hepatocellular Carcinoma](#)
- [Towards better healthcare: What could and should be automated?](#)

- [A Deep Multi-Modal Method for Patient Wound Healing Assessment](#)
- [Evolution-based Fine-tuning of CNNs for Prostate Cancer Detection](#)
- [HR-CAM: Precise Localization of Pathology Using Multi-level Learning in CNNs](#)
- [Deep learning feature based medical image retrieval for large-scale datasets](#)

8 Considerações finais

De maneira geral, temas que foram recorrentes na conferência e são oportunidades para trabalhar na área:

- Melhorar/incluir interpretabilidade e explicabilidade nos modelos
- Bias é um problema sério e devemos desenvolver métodos para lidar com isso
- Privacidade do usuário está em alta com a nova área de Federated Learning
- Deep Reinforcement Learning está muito em alta principalmente por aplicações em jogos
- GANs continuam em alta, mas perdeu força em relação aos conferências anteriores
- Neurociência também está em alta, principalmente para prover teoria e ideias para desenvolver modelos melhores dos que a área possui

Além dos tópicos mencionado, vale a pena notar que ética em AI é algo que boa parte da comunidade está bastante preocupada. Palestras e apresentações frisavam que AI não deve ser vista como algo que vai substituir pessoas e sim trabalhar colaborativamente. Além disso, discussões foram levantadas a todo momento em relação até onde devemos aplicar AI. Por exemplo, dois artigos que geraram muita controvérsia e repercussão na comunidade: [Face Reconstruction from Voice using Generative Adversarial Networks](#) e [Predicting the Politics of an Image Using Webly Supervised Data](#). Muitos taxaram esses artigos com racistas e outros como inúteis, no sentido de que não agrega em nada para a sociedade. De fato é algo que nossa comunidade tem que pensar sério, aprender e evoluir.

Para concluir, a conferência é extremamente cansativa, mas vale muito a pena participar. O que eu vi foi que eles estão preocupados com inclusão e em prover uma comunidade consciente do papel da AI no mundo. Como as palestras principais frisaram, nossa área afeta a vida de bilhões de pessoas diariamente e não existe algoritmo neutro. Embora ocorram falhas, como os dois artigos citados no parágrafo anterior, acredito que a conferência tem evoluído para o caminho correto.