Conclusion and Future Research Directions

Recent Advances in Transferable Representation Learning (Part IV)

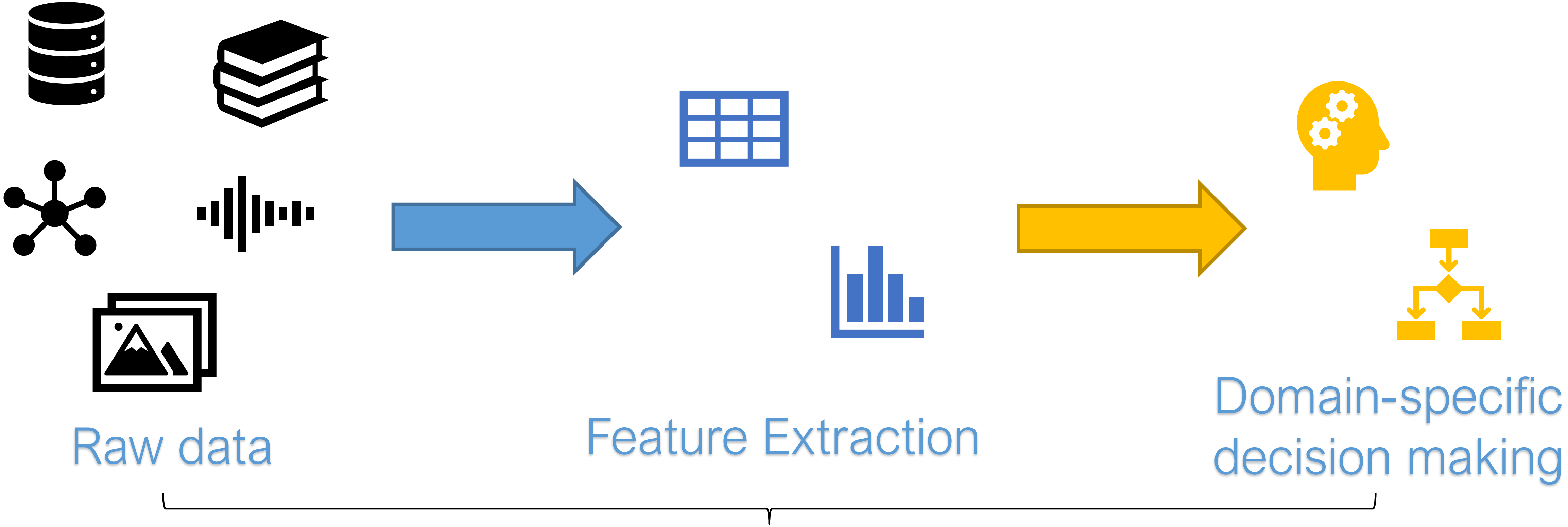
Muhao Chen, Kai-Wei Chang, Dan Roth

Feb 2020

AAAI Tutorials

Recent Advances in Transferable Representation Learning

What has representation learning enabled?



Raw data

Feature Extraction

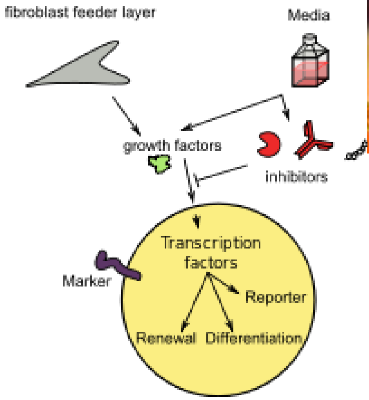
Domain-specific decision making

Automatic, end-to-end

Why Transferability is Important

- In some domains, we have lots of learning resources.
- In other domains, learning resources are insufficient.

漢語
汉语
中文



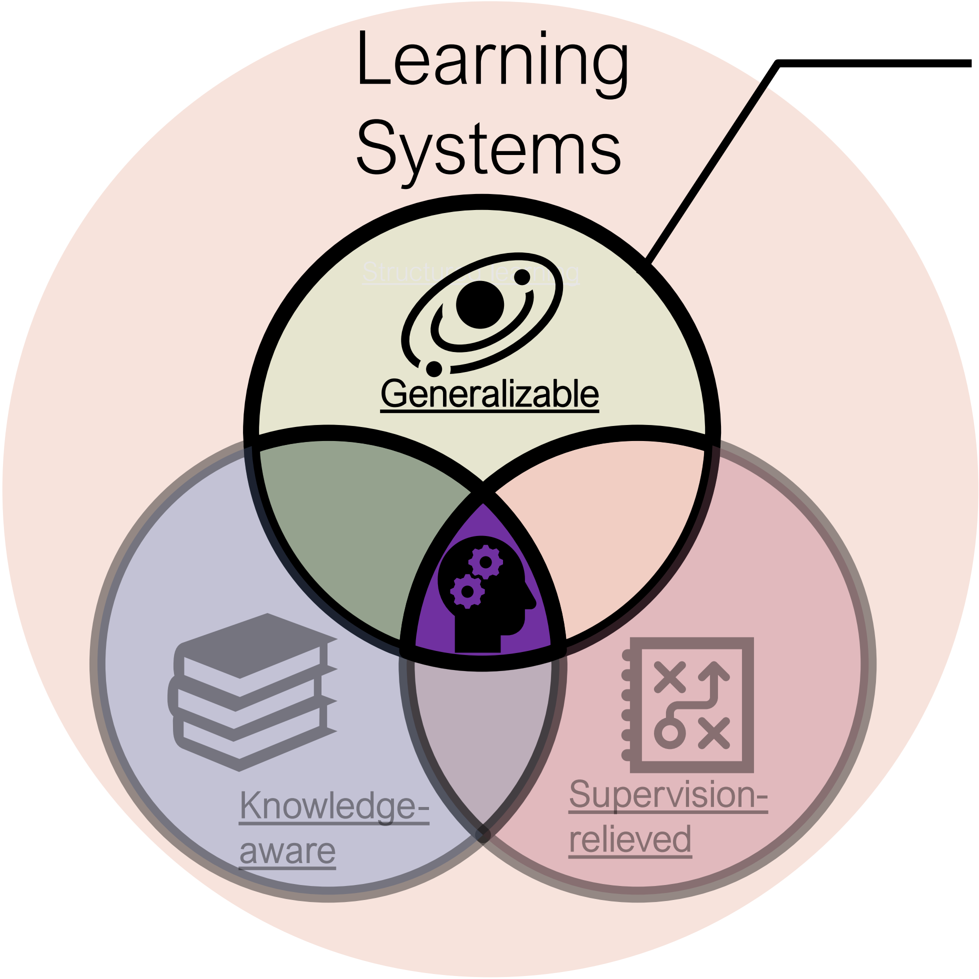
High-resource domains

التاميل

عربى

Low-resource domains

Why Transferability is Important



- Knowledge is interchangeable across different domains.
- Leveraging the knowledge from high-resource domains to help decision making in low-resource domains.
- Making learning and inference **generalizable** and **adaptive**.

Conclusion

- Research Questions We Have Discussed
  - □ Languages
    - Can we learn representations of concepts in a way that is independent of the language?
    - Can we use it to perform well in languages with very little annotated data?
  - □ Modality
    - Can we learn representations that capture both visual and textual properties?
    - Can we use it to improve performance on relevant tasks?
  - □ Domains
    - Can we capture the association of knowledge with limited supervision?
    - Can we effectively populate missing knowledge in domains?

Several Perspectives of Future Work

- Transferable representations for highly complex structures
  - □ Hierarchical structures
  - □ Order-invariant structures
- Fairness and trustworthiness in knowledge transfer
- The emerging application scenarios requiring transferable representation learning

Transferable hyperbolic representation learning

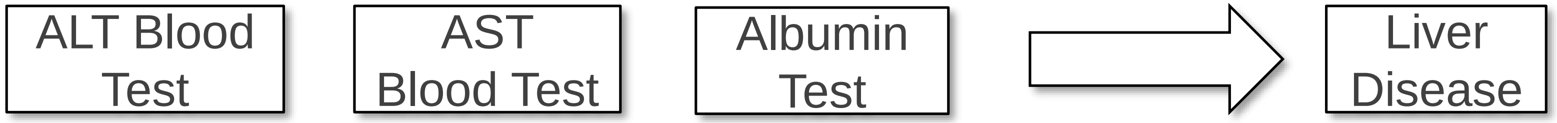
- Many data form hierarchies
  - Ontologies, taxonomies, syntax trees, org charts, citation graphs, etc.

- Particularly suitable for a **hyperbolic space**
  - The amount of space increases exponentially w.r.t. the radius [Nickel+ NIPS-17, Ganea+ NeurIPS-18, Liu+ NeurIPS-19]

- Transferable hyperbolic representation learning benefits tasks
  - Ontology matching and population
  - Label space transfer for hierarchical classification
  - Transfer learning on programming languages (or ASTs)

Transferable set learning

- Unordered and unsized data (i.e. forming a *set*)
  - Point cloud
  - Clinical events in single-visit electronic health records (EHR)
- Set learning: order-invariant representation learning
  - Differentiable pooling [Zaheer+ NIPS-17]
  - Permutation neural networks [Meng+ KDD-19]
- Applications
  - Risk prediction on EHR data: given a set of lab tests, predict possible diseases / future clinical events



ALT Blood Test

AST Blood Test

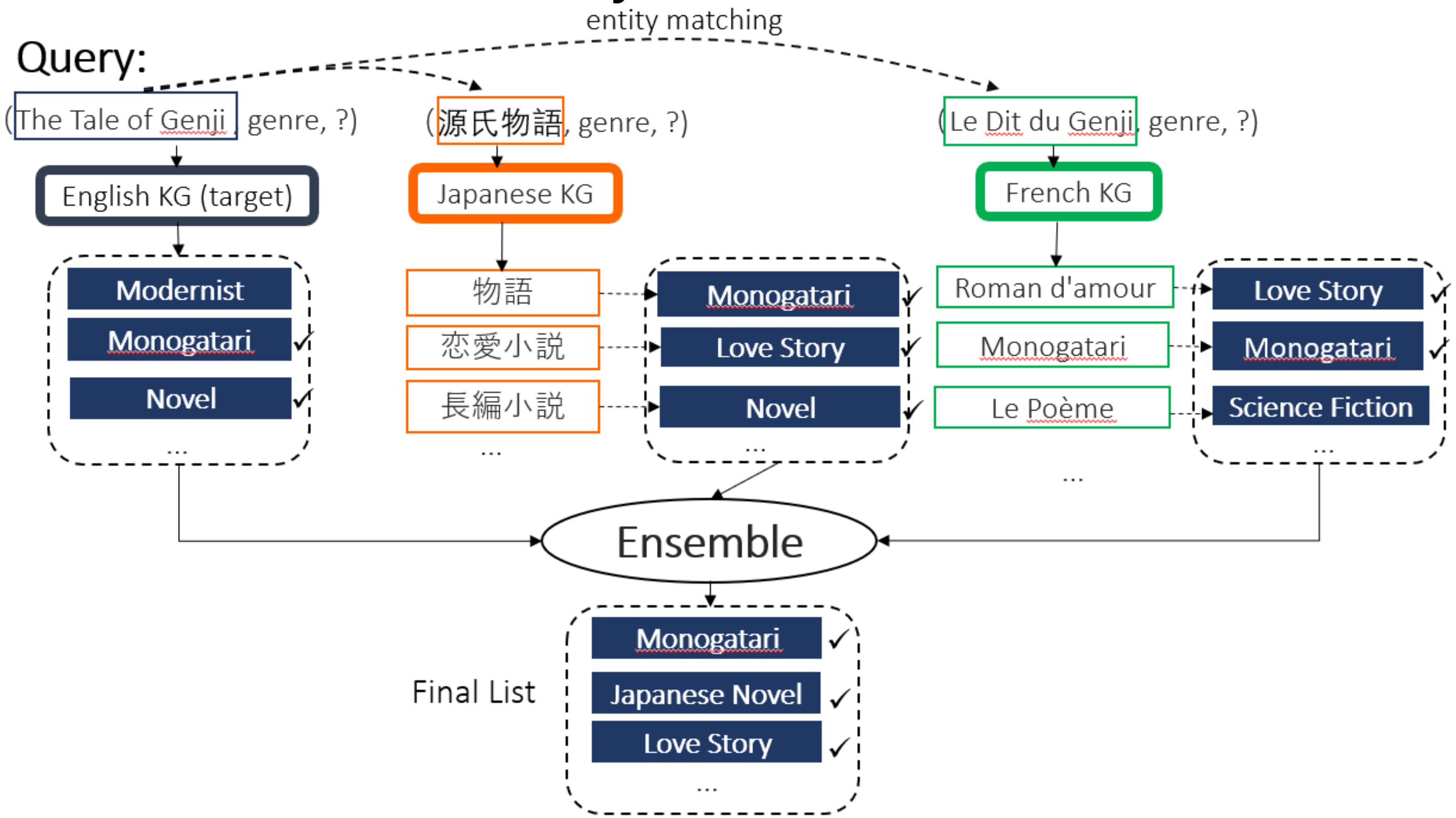
Albumin Test

Liver Disease

  - Self-driving: learning from a sensor point cloud to predict driving actions
- Why transferability
  - Clinical data are often low-resource due to privacy
  - Models must be generalizable in clinical and self-driving scenarios

Fairness and Trustworthiness

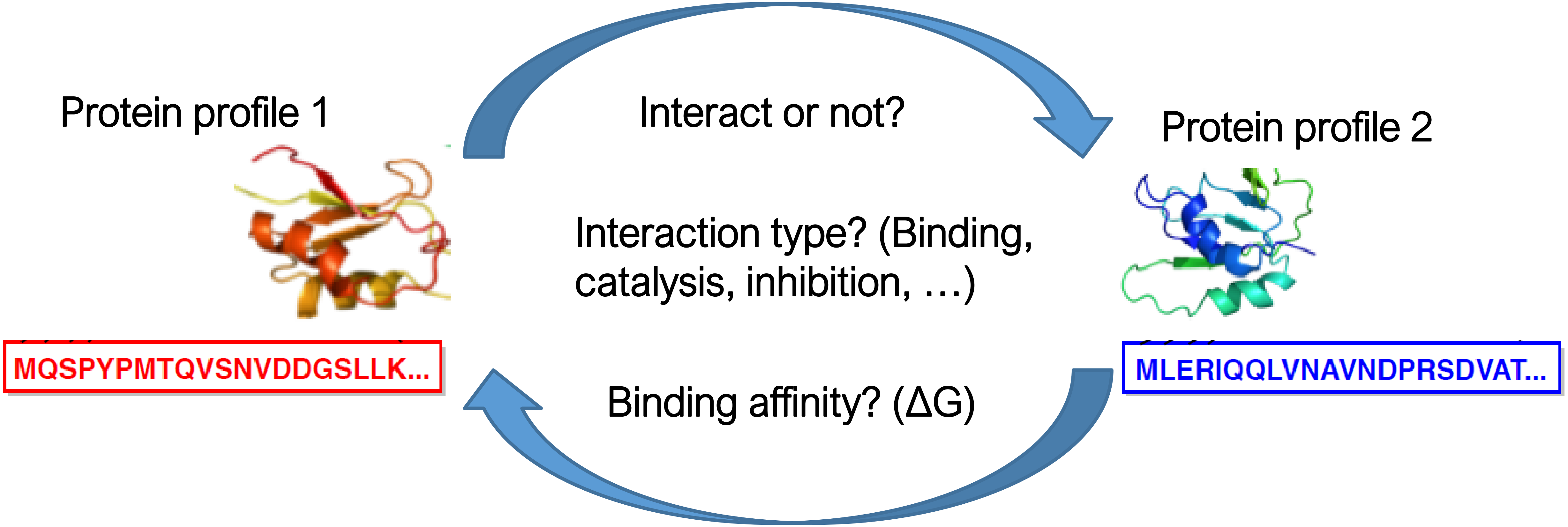
- Trustworthiness: when combining multiple sources of knowledge, which one should we believe when there is inconsistency?



- Fairness: how do we mitigating societal bias in different domain/language-specific data?

An Emerging Area: Representation Learning for Genomic Data

*[Chen+ *ISMB'19, Bioinformatics 2019*]



Protein profile 1

MQSPYPMTQVSNVDDGSLLK...

Interact or not?

Interaction type? (Binding, catalysis, inhibition, …)

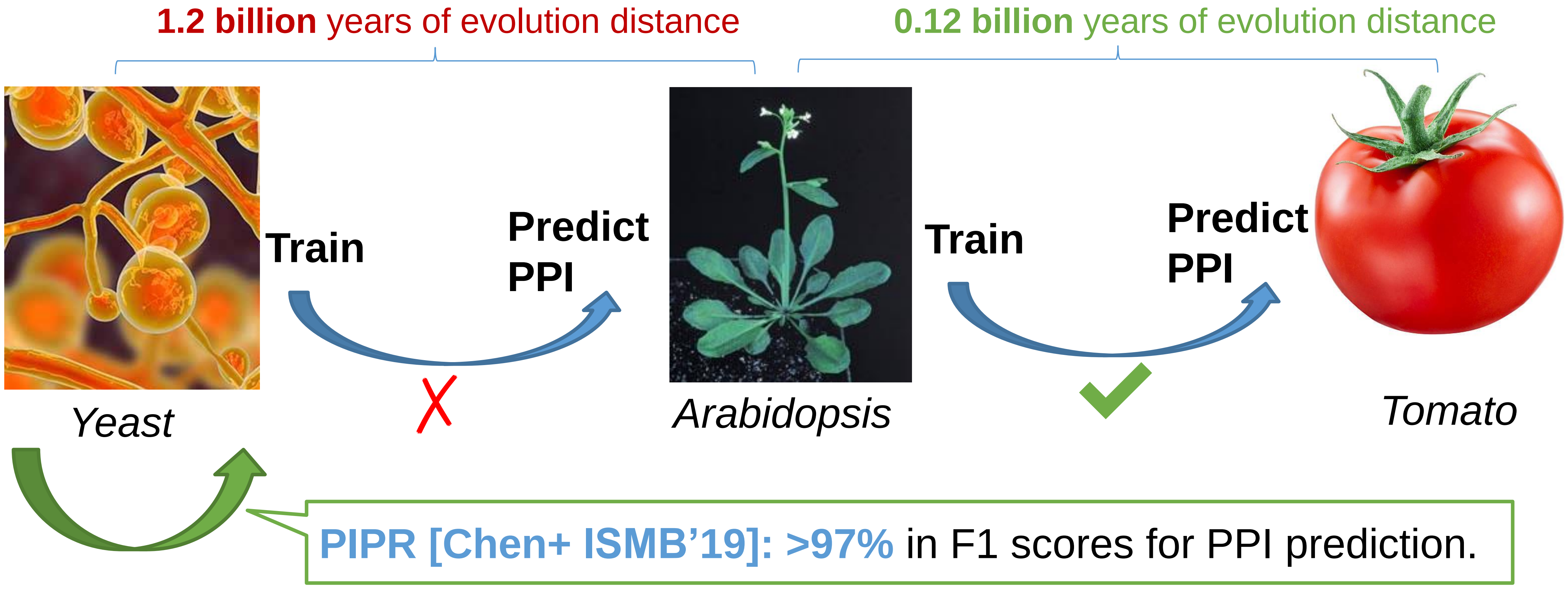
Binding affinity? (ΔG)

Protein profile 2

MLERIQQLVNAVNDPRSDVAT...

An example task: Protein-protein interaction prediction.

Cross-species Transferability: Why Important



1.2 billion years of evolution distance

0.12 billion years of evolution distance

Yeast

Train

Predict PPI

✗

Arabidopsis

Train

Predict PPI

✓

Tomato

PIPR [Chen+ ISMB’19]: >97% in F1 scores for PPI prediction.

Emerging topic: transferability across species

Cross-species Transferability: What Are Needed?

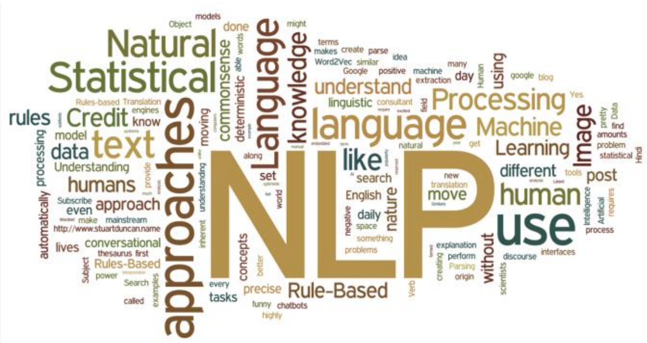
- **<3.5k** “high-resource” species vs **1.5M** “low-resource” species
  - □ Complex organisms without full genomes
  - □ Newly discovered ones

- Transferred learning is important for *de novo* prediction on **1.5M** “low-resource” species
  - □ Reliable *de novo* prediction can be used to guide wet lab experiments

- New technologies for the community
  - □ Adversarial learning for **“species-invariant” sequence representations**
  - □ Massively **pre-trained language models** for amino acid sequences

Cross-domain and Interdisciplinary Research

Transferable representation learning could address problems in **multiple research areas**.
There are lots of challenges before making it **work for Good**.



Natural Language Processing

QA & Dialogue systems

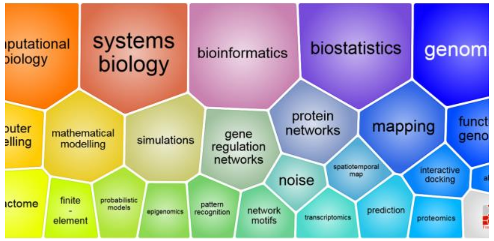
Knowledge Representation

ebay

amazon

NETFLIX

Recommendation Systems



Computational Biology

Thank You