Recent Advances in Transferable Representation Learning

Muhao Chen, Kai-Wei Chang, Dan Roth

AAAI 2020 Tutorial

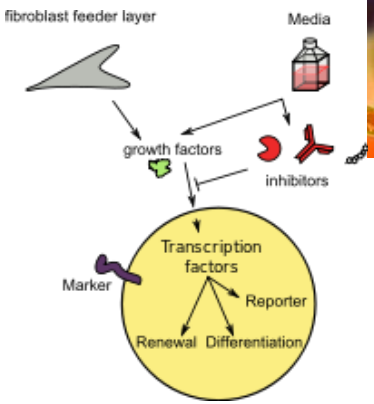
Transfer

- Transfer is one of the key challenges to AI
- Given a classification problem, and enough data, we know that we can learn a model that will do well on “similar” data.
  - Learning theory has supported good understanding of these situations for years
- But, what about not-so-similar data?
  - The notions of **transfer** and **adaption** have been persistent key challenges to machine learning
- Transfer across domains, tasks, languages, modalities

Why Transferability is Important

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

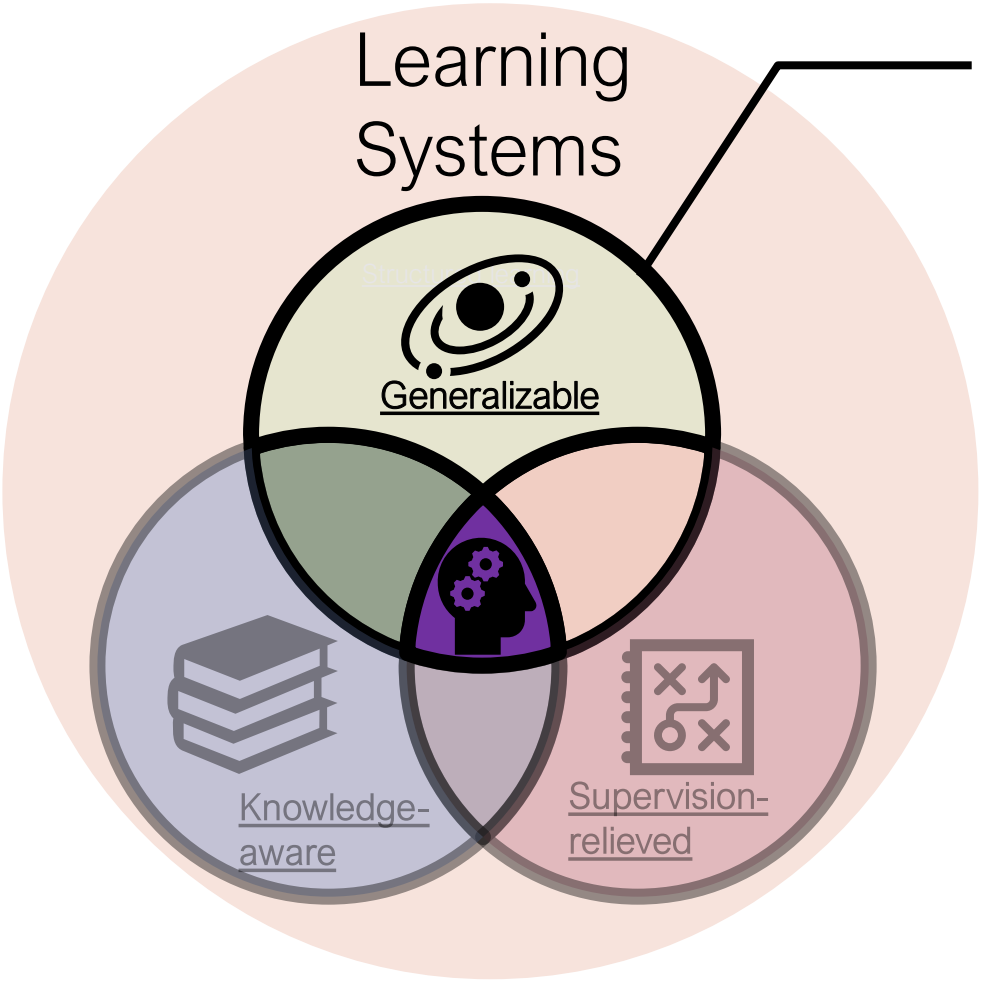
漢語
汉语
中文



High-resource domains

Low-resource domains

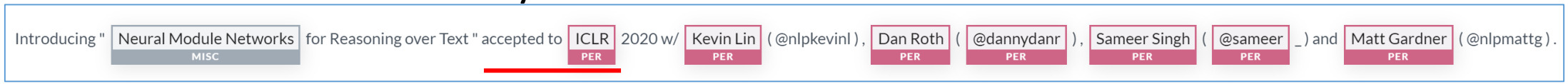
Why Transferability is Important



- 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**.

Natural Language Processing

- Adaptation is essential in NLP.
- Vocabulary differs across domains
  - Word occurrence may differ, word usage may differ; word meaning may be different.
  - "can" is never used as a noun in a large collection of WSJ articles
- Structure of sentences may differ



- Task definition may differ

State-of-the-art Named Entity Recognition

Model	CoNLL	OntoNotes	Enron (email)
Flair trained on CoNLL	93.03	76.82 ↓	56.80 ↓

Domain Adaptation

Semantic Role Labeling: Shallow Semantic Parsing

UN Peacekeepers abuse children

Semantic Role Labeling Demo

Input Text:

UN peacekeepers abuse children.

Click For General Explanation of Argument Labels

Output:

SRL

UN	agent [A0]
peacekeepers	V: peacekeepers
abuse	patient [A1]
children	
.	

Wrong!
"Peacekeepers" is not the Verb

Reason: "abuse" was never observed in training as a verb

All state-of-the-art systems, including those based on BERT make this mistake.

UN Peacekeepers hurt children

Semantic Role Labeling Demo

Input Text:

UN peacekeepers hurt children.

Correct!

Click For General Explanation of Argument Labels

Output:

SRL

UN	agent, entity causing damage [A0]
peacekeepers	
hurt	V: hurt
children	patient, entity experiencing hurt/damage [A1]
.	

Example from Kundu & Roth, CoNLL'11: Adapting Text Instead of the Model: An Open Domain Approach

What Has Happened?

- Nevertheless, we have seen significant progress in our ability to transfer representations.

- We will focus on methods, and exemplify it in three domains

  - Cross-lingual Transfer
    - Can we train models on one language and use it successfully in another?
  - Transfer across domains
    - Can we transfer knowledge from one domain to help in other, related domains?
    - Gene ontologies → Protein-Protein Interaction
  - Transfer in a multi-modal setting
    - Can we use information learned from text to support visual inferences?

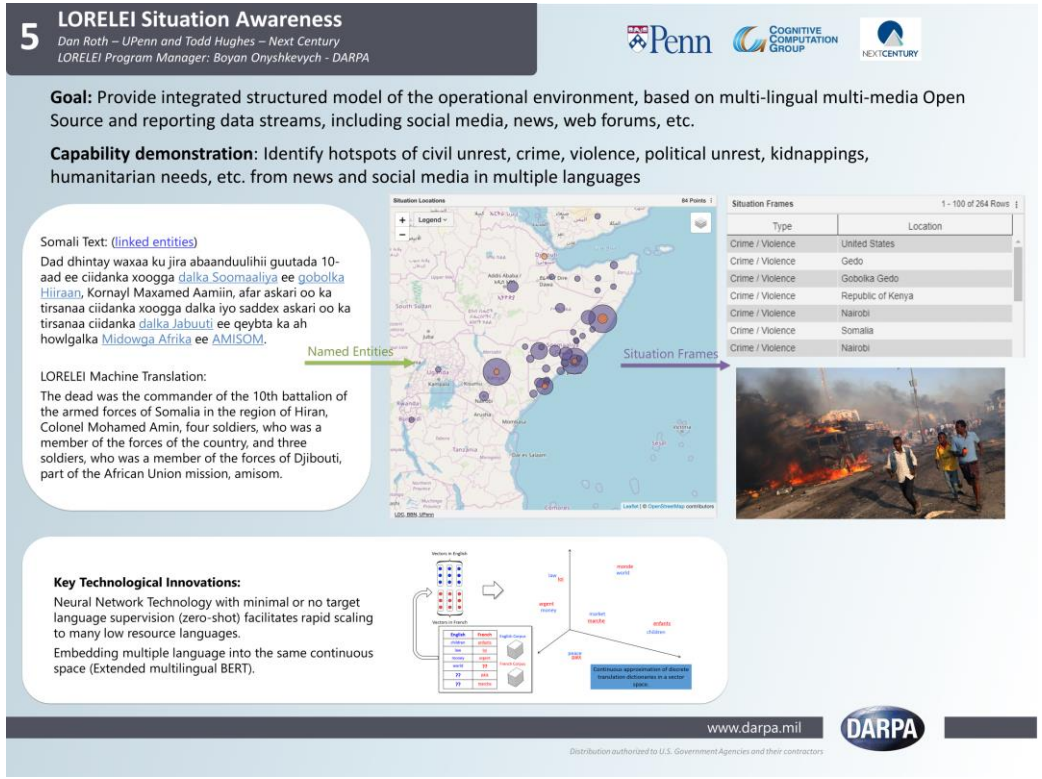
Cross-Lingual Natural Language Processing

- Goal: Given text data in a low resource language,
  - Can we "understand" it even if you only know English?
  - **No training data** in the low resource language!

Somali streaming data → Situation Awareness (described in English)

Goal

- "Understand" a situation described in Target Language
  - Identify Entities & Concepts (NER)
  - Ground in English Resources (EDL)
  - Type the situation



Cross-lingual Entity Linking (XEL) [Tsai & Roth NAACL' 16, Upadhyay et al. EMNLP'18]

Given a (non-English) document, extract named entities and disambiguate into the English Wikipedia (KB)

… its lead singer Nunn left Berlin to audition for Star Wars …

… sein Leadsänger Nunn verließ Berlin, um für Star Wars vorzuspielen …

… 其主唱 纳恩 离开 柏林 去参加 星球大战 的试镜 …

Terri Nunn

From Wikipedia, the free encyclopedia

Terri Kathleen Nunn (born June 26, 1961[1]), is an American singer and actress. She is best known as the lead vocalist of the new wave/synthpop band Berlin.

Contents [hide]
1 Biography
1.1 Personal life
2 References
3 External links

Terri Nunn performing at Oracle OpenWorld 2010.

Berlin (band)

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(August 2013) (Learn how and when to remove this template message)*

Berlin is an American new wave band. The group was formed in Orange County in 1978 by John Crawford (bass guitar). Band members included Crawford, Terri Nunn (vocals), David Diamond (keyboards), Ric Olsen (guitar), Matt Reid (keyboards) and Rod Learned (drums). The band gained mainstream-commercial success in the early 1980s with singles including "The Metro", "Sex (I'm A...)", "No More Words" and then in

Star Wars

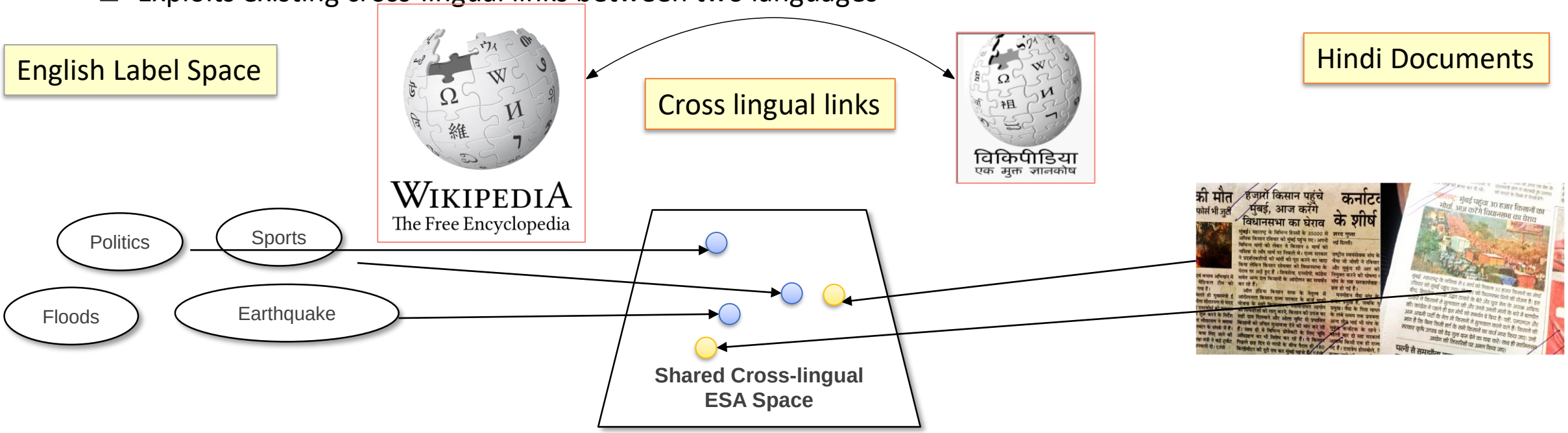
From Wikipedia, the free encyclopedia

This article is about the film series and media franchise. For the original 1977 film, see Star Wars (film). For other uses, see Star Wars (disambiguation).

Star Wars is an American epic space opera media franchise, centered on a film series created by George Lucas. It depicts the adventures of various characters "a long time ago in a galaxy far, far away".

Cross-lingual Textual Classification [Song et al. IJCAI'16]

- Text Classification with No Annotated data
- A cross-lingual representation: Cross-lingual Explicit Semantic Analysis (CL-ESA)
  - Exploits existing cross-lingual links between two languages

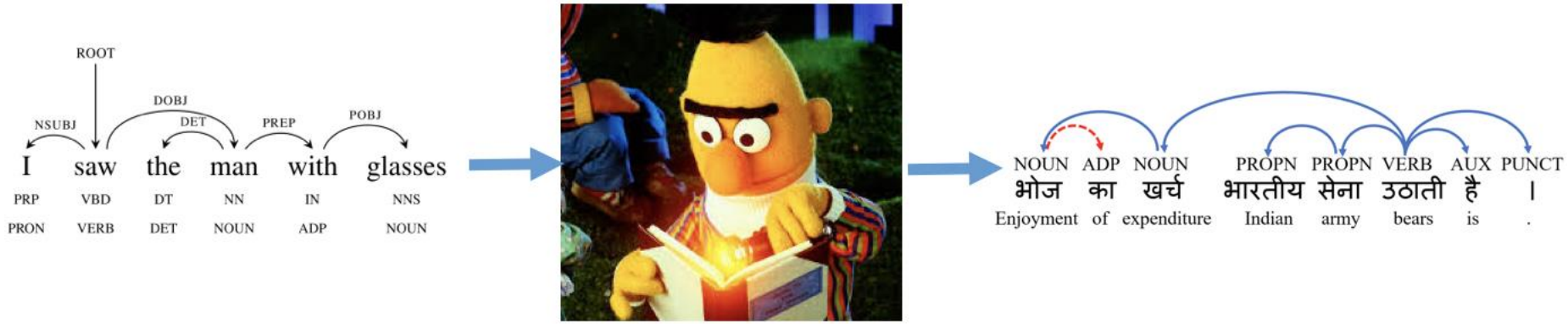


  - Represent low-resource language documents and English Labels in the corresponding CL-ESA Space
  - Map representation via cross-lingual links.
  - Quality depends on the (size of the) intersection of the title spaces

Many Additional Tasks

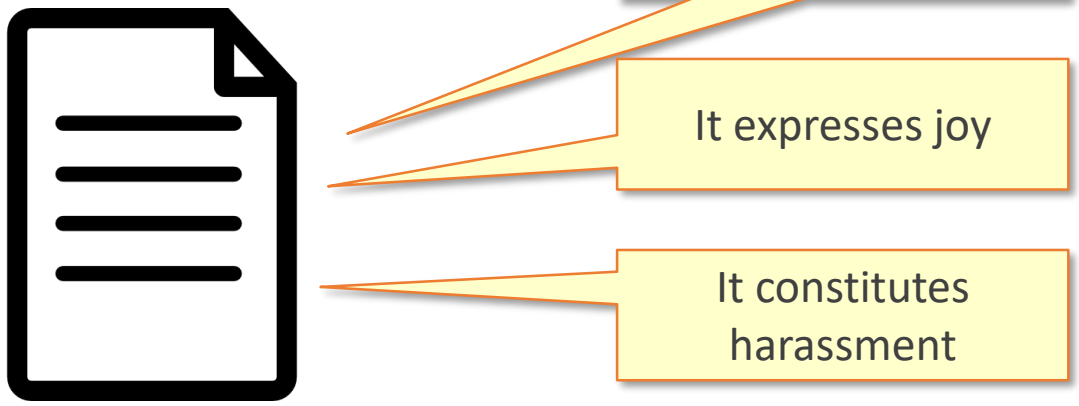
- For many tasks, we have training data in English, but not in other languages

- Can we train a model for more difficult tasks, (e.g., parsing) in English and transfer it to other language?
- This ability often depends on assumptions we can make:
  - E.g., that languages are aligned in a (contextualized) word embedding space
  - Then, we can model the commonality between languages, and inject information about target language in inference



Task Transfer: From Textual Entailment to 0-Shot-Text Classification

- Text Classification is inherently a Textual Entailment problem.
  - But we don't think about it when we have a lot of labeled data
- Given a document or a text snippet:
  - Classifying it is equivalent to determining the truth value of a hypothesis, given the text snippet
- A program that "knows" how to determine textual entailment can thus classify with respect to any label (hypothesis).



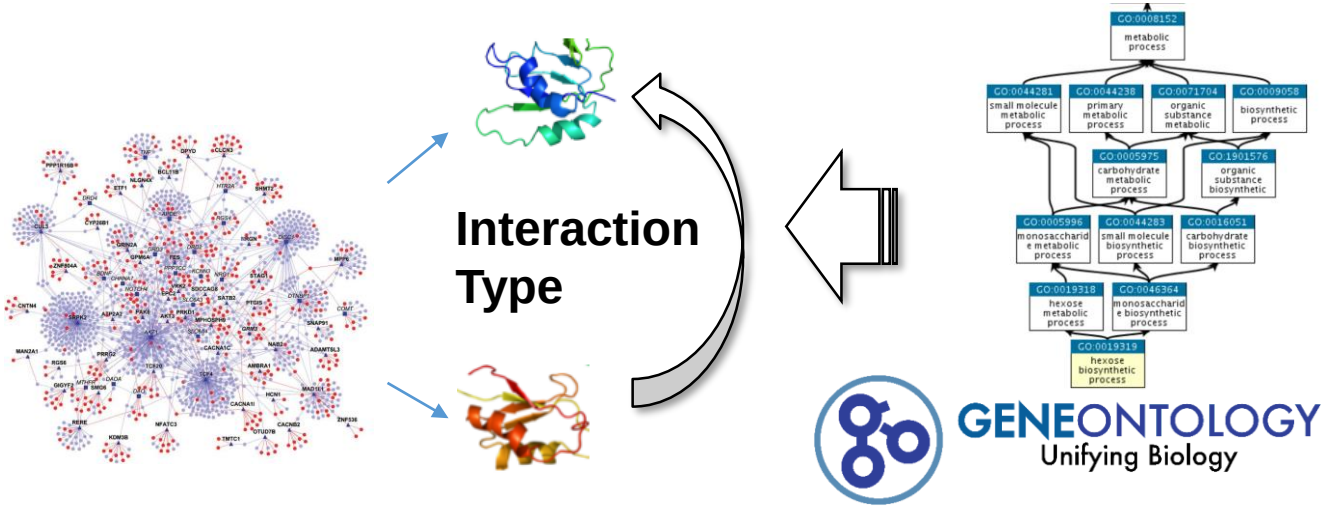
- Assuming, of course, that it "understands" the meaning of the hypothesis.

1. A model that "knows" the trick of determining textual entailment – independent of the task at hand

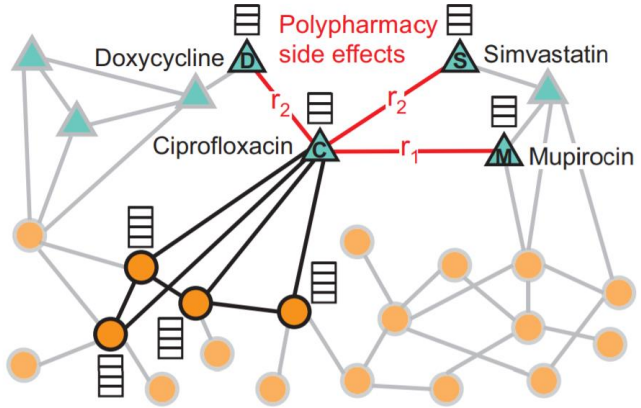
2. A model that "understands" the label – part of the input.

Transfer in Semi-Structured Domains

- Given a graph, predict the edge type (e.g. protein-protein interactions, drug interactions)
- Different graphs can be aligned
- Additional graphs provide more context information of nodes, and help inferring the edge types
  - Multiple knowledge bases
  - Protein networks and gene ontologies
  - Drug networks and protein networks



Protein-protein Interaction



Drug Interaction

Transfer in a Multi-Modality Setting

- Can we transfer learned information across-modalities?
  - A joint representation space that captures the information across different modalities
  - Associations between modalities allow transfer of knowledge to downstream tasks

Is it raining outside?

a) Yes, it is snowing.
b) Yes, `[person8]` and `[person10]` are outside.
c) No, it looks to be fall.
d) Yes, it is raining heavily.

An example from the VCR dataset

Transfer to answering commonsense questions

Intuition

- It has always been clear that a key notion in facilitating transfer is that of abstraction

- Can we **abstract away** from properties of the domain or language, and represent the **essential properties of the problem**?

  - UN peacekeepers abuse children → UN peacekeepers hurt children

- A lot of efforts attempted to do just that in Feature Spaces

Intuition (2)

- Ben David et al., Machine Learning Journal, 2010:
  - To achieve domain adaptation one needs to find a representation that *explicitly minimizes the difference between the source and target domains, while at the same time maximizing the margin of the training set.*
- This was the key insight also in several experimental frameworks:
  - Daume: Frustratingly easy domain adaptation
  - Blitzer et al.: structural correspondence learning
  - …..

Representation

- In recent years, the community has developed better understanding of representations.
- In many (but not all) cases, these are continuous representations,
  - And they seem to provide better abstractions of the data than earlier representations we used in machine learning.
- The key questions this tutorial addresses:
  - Can we utilize this technology to facilitate better transfer?
  - How?

Types of Transfer

- We will address these questions from three distinct perspectives:
  - Languages
    - Can we learn representation of concepts in a way that is independent of the language they are expressed in?
    - 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 learn representations that abstract key domain properties?
    - Can we effectively transfer knowledge across domains?

Tutorial Outline

- Introduction 2:00 – 2:30
  - Dan Roth
- The basics of embeddings and Cross-X embeddings 2:30 – 2:45
  - Kai-Wei Chang
- Cross-Lingual Representations and Transfer 2:45 – 3:45
  - Dan Roth, Kai-Wei Chang
- Break 3:45 – 4:15
- Multimodal Representations and Transfer 4:15 – 4:45
  - Kai-Wei Chang
- Cross-domain Transfer for Multi-relational Representations 4:45 – 5:45
  - Muhao Chen
- Conclusion and Future Work 5:45 – 6:00
  - Muhao Chen