AIHL
AI for Healthy Living

UC San Diego IBM

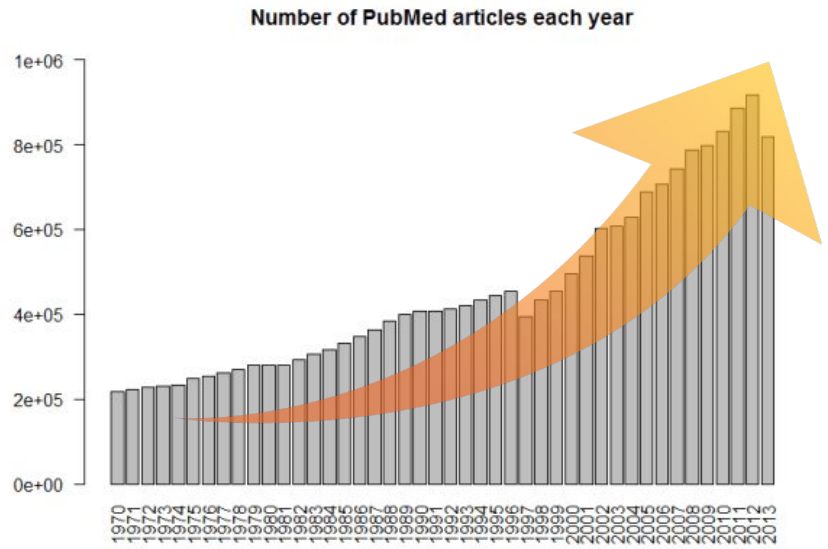
# Abstractified Multi-instance Learning (AMIL) for Biomedical Relation Extraction

Bill Hogan[1,2], Molly Huang[2], Yannis Katsis[4],
Tyler Baldwin[4], Ho-Cheol Kim[4], Yoshiki Vazquez-Baeza[2],
Andrew Bartko[2], Chun-Nan Hsu[3]

This work is supported by IBM Research AI through the AI Horizons Network

*[1]Department of Computer Science & Engineering, [2]Center for Microbiome Innovation, [3]Center for Research in Biological Systems*
*University of California, San Diego, La Jolla, CA 92093*
*[4]IBM Research-Almaden, 650 Harry Road, San Jose, CA 95120*

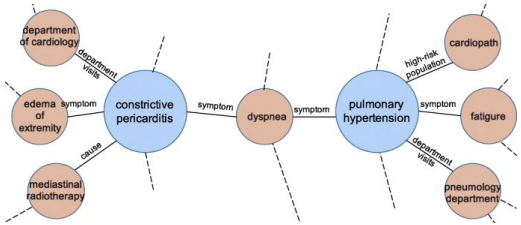
# How to automatically extract fact triples (h, r, t) from large amounts of raw biomedical text (e.g. PubMed)?




https://pubmed.ncbi.nlm.nih.gov/

# Background: Distant Supervision

## Knowledge Graph (UMLS[1])




9.9M fact triples

+

## Raw Text (PubMed[2])

The Human Microbiome and Obesity: Moving beyond Associations

Padma Maruvada [1], Vanessa Leone [2], Lee M Kaplan [3], Eugene B Chang [4]

Affiliations + expand

PMID: 29120742 DOI: 10.1016/j.chom.2017.10.005

Abstract

Mounting evidence indicates that the gut microbiome responds to diet, antibiotics, and other external stimuli with speed and high precision and in ways that impact a variety of metabolic conditions including obesity and non-alcoholic fatty liver disease. Despite a decade of research establishing a strong association between the gut microbiota and obesity in humans, a causal relationship and the underlying mechanism remain outstanding. Several technological and methodological limitations in obesity and microbiome research have made it difficult to establish causality in this complex relationship. Additionally, limited collaborative interaction between

32M+ articles

**r? (tibia, fibula)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

}

**! Very noisy labels**

**Auto-generated label:**
*articulates_with*

[1]https://www.nlm.nih.gov/research/umls/index.html, [2]https://pubmed.ncbi.nlm.nih.gov/

# Background: Multi-instance Learning

## Single Instance Learning

**r? (tibia, fibula)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

**r? (tibia, fibula)**

The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

**r? (tibia, fibula)**

The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

## Multi-instance Learning

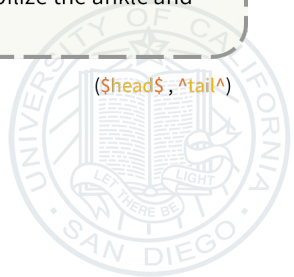
**r? (tibia, fibula)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

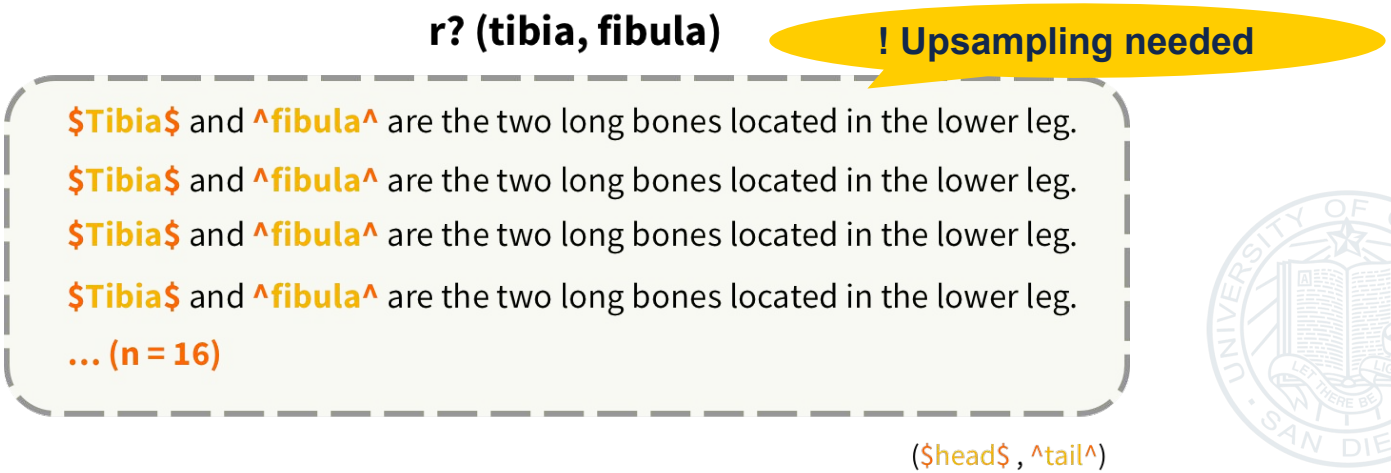
The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

($head$ , ^tail^)



# Background: Multi-instance Learning

- After preprocessing PubMed, ~52% entity pairs are support by one to two sentences.
- Upsampling used to fill bags when less than 16 sentences are available.



# Our Contribution: Abstractified Multi-instance Learning (AMIL)

## Single Instance Learning

**r? (tibia, fibula)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

**r? (tibia, fibula)**

The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

**r? (humerus, ulna)**

It is composed of two bones, the ^humerus^ and $ulna$, and is the junction between the trochlear notch of ulna and the trochlea of humerus.

**r? (humerus, ulna)**

The accurate adaptation of the trochlea of the ^humerus^, with its prominences and depressions, to the trochlear notch of the $ulna$, prevents any lateral movement.

## Multi-Instance Learning

**r? (tibia, fibula)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

**r? (humerus, ulna)**

It is composed of two bones, the ^humerus^ and $ulna$, and is the junction between the trochlear notch of ulna and the trochlea of humerus.

The accurate adaptation of the trochlea of the ^humerus^, with its prominences and depressions, to the trochlear notch of the $ulna$, prevents any lateral movement.

## Abstractified Multi-instance Learning

**Bag: ($body part$, ^body part^)**

$Tibia$ and ^fibula^ are the two long bones located in the lower leg.

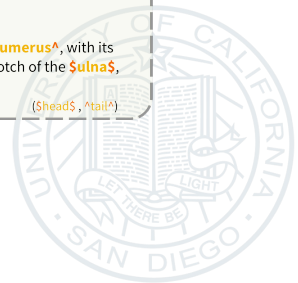
The ^fibula^ supports the $tibia$ and helps stabilize the ankle and lower leg muscles.

It is composed of two bones, the ^humerus^ and $ulna$, and is the junction between the trochlear notch of ulna and the trochlea of humerus.

The accurate adaptation of the trochlea of the ^humerus^, with its prominences and depressions, to the trochlear notch of the $ulna$, prevents any lateral movement.

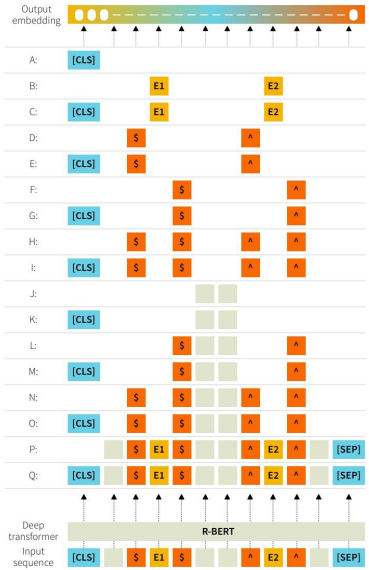
($head$ , ^tail^)

**r? (tibia, fibula)**



# Relationship Embedding Experiments

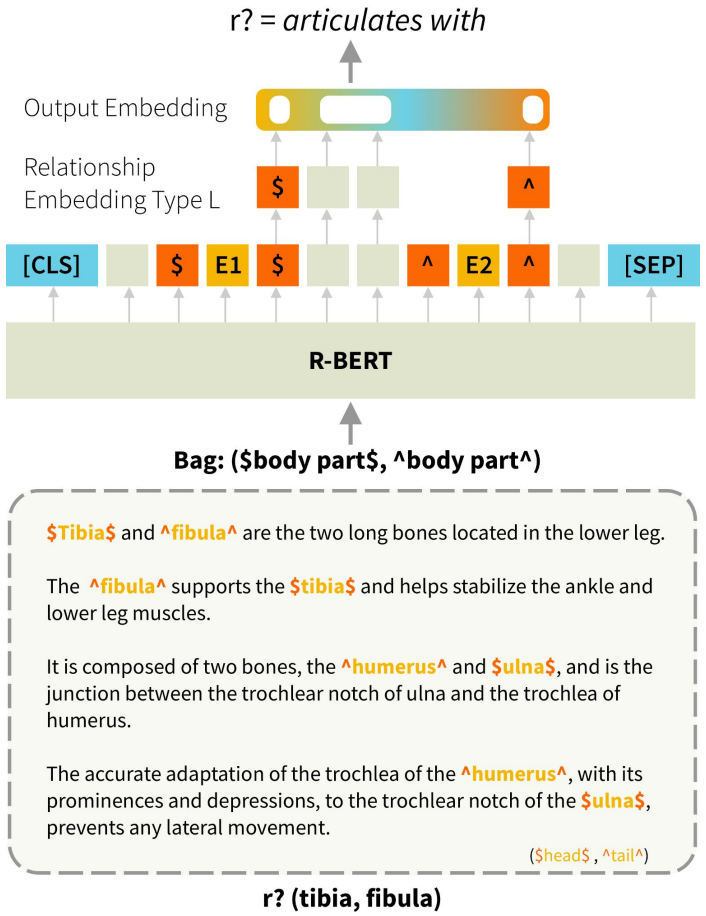
| | |
|---|---|
| A: | [CLS] |
| B: | entity mention pool |
| C: | [CLS] + entity mention pool |
| D: | $e_{1Start} + e_{2Start}$ |
| E: | [CLS] + $e_{1Start} + e_{2Start}$ |
| F: | $e_{1End} + e_{2End}$ |
| G: | [CLS] + $e_{1End} + e_{2End}$ |
| H: | $e_{1Start} + e_{1End} + e_{2Start} + e_{2End}$ |
| I: | [CLS] + $e_{1Start} + e_{1End} + e_{2Start} + e_{2End}$ |
| J: | middle mention p. |
| K: | [CLS] + middle mention p. |
| L: | $e_{1End}$ + middle mention pool + $e_{2End}$ |
| M: | [CLS] + $e_{1End}$ + middle mention p. + $e_{2End}$ |
| N: | $e_{1Start} + e_{1End}$ + middle mention p. + $e_{1End} + e_{2End}$ |
| O: | [CLS] + $e_{1Start} + e_{1End}$ + middle mention p. + $e_{1End} + e_{2End}$ |
| P: | entire sequence avg |
| Q: | [CLS] + entire sequence avg |



AIHL
AI for Healthy Living

UC San Diego IBM

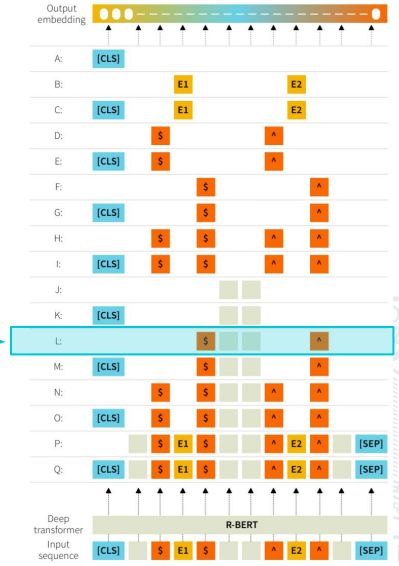
# AMIL Model



| RE Model | AUC | F1 | P@2k | P@4k | P@6k | P@10k | P@20k |
|---|---|---|---|---|---|---|---|
| Dai et al. (2019)* | — | — | .913 | .829 | .753 | — | — |
| Amin et al. (2020)* | .684 | .649 | .983 | .977 | .969 | — | — |
| Amin et al. (2020) w/Sci-Spacy | .758 | .710 | .998 | .995 | .991 | .981 | .905 |
| AMIL | .862 | .795 | .997 | .997 | .997 | .994 | .947 |
| AMIL Rel. Type 'L' | **.872** | **.812** | **1.000** | **.999** | **.999** | **.995** | **.953** |

| | Relationship Representation | F1 | AUC | P@20k |
|---|---|---|---|---|
| A: | [CLS] | .793 | .863 | .947 |
| B: | entity mention pool | .786 | .855 | .943 |
| C: | [CLS] + entity mention pool | .795 | .862 | .947 |
| D: | $e_{1Start} + e_{2Start}$ | .795 | .859 | .948 |
| E: | [CLS] + $e_{1Start} + e_{2Start}$ | .792 | .860 | .946 |
| F: | $e_{1End} + e_{2End}$ | .804 | .872 | .951 |
| G: | [CLS] + $e_{1End} + e_{2End}$ | .799 | .861 | .950 |
| H: | $e_{1Start} + e_{1End} + e_{2Start} + e_{2End}$ | .792 | .857 | .947 |
| I: | [CLS] + $e_{1Start} + e_{1End} + e_{2Start} + e_{2End}$ | .780 | .859 | .949 |
| J: | middle mention p. | .805 | .862 | .952 |
| K: | [CLS] + middle mention p. | .788 | .850 | .945 |
| L: | $e_{1End}$ + middle mention pool + $e_{2End}$ | **.812** | **.872** | **.953** |
| M: | [CLS] + $e_{1End}$ + middle mention p. + $e_{2End}$ | .804 | .865 | .951 |
| N: | $e_{1Start} + e_{1End}$ + middle mention p. + $e_{1End} + e_{2End}$ | .800 | .865 | .950 |
| O: | [CLS] + $e_{1Start} + e_{1End}$ + middle mention p. + $e_{1End} + e_{2End}$ | .804 | .865 | .950 |
| P: | entire sequence avg | .800 | .862 | .948 |
| Q: | [CLS] + entire sequence avg | .808 | .864 | .949 |

Best relationship representation



# Thank You