Improved Semantic Representation for Domain-Specific Entities

Mohammad Taher Pilehvar and Nigel Collier

12 August 2016, BioNLP
Semantic Representation

Representing meaning in a machine-interpretable (mathematical) form
Semantic Representation
Vector Space Model

![Diagram of semantic representation with vectors for pizza, restaurant, food, and bank.](image)
Semantic Representation of Phenotypes

“... The patient was reported to have headache, nausea, vomiting, abdominal pain, weakness, increased blood pressure, and skin rashes. Sequential therapy has been reported to have a lower incidence of side effects but a higher rate of pregnancy. ...”

Abnormal observed properties of an organism, such as abnormal morphology/structure/behavior
Representation of phenotypes
Applications: Disease phenotype relations
Representation of phenotypes
Applications: Differential diagnosis

- Chest Pain
- Ebola Hemorrhagic Fever
- Gingerval Bleeding
- Skin Rash
- Migraine
- Typhoid
- Hallucinations
Representation of phenotypes

Applications: Cross-species linking of lab results
> Translational medicine
Representation of phenotypes

Difficulties

• Domain specific
  • (hence, usually rarely used in general-domain corpora - low frequency)

• Usually multi-words: compositionality

• Fine granularity
Representation of phenotypes
Difficulties: Rare or out-of-vocabulary words

Artificial intelligence (AI) is the intelligence exhibited by machines or software. It is also the name of the academic field of study which studies how to create computers and computer software that are capable of intelligent behavior. Major AI researchers and textbook define this field as “the study and design of intelligent agents”\(^2\) in which an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success.\(^3\) John McCarthy, who coined the term in 1955,\(^4\) defined it as “the science and engineering of making intelligent machines”\(^5\).

AI research is highly technical and specialized, and is deeply divided into subfields that often fail to communicate with each other.\(^6\) Some of the division is due to social and cultural factors; subfields have grown up around particular institutions and the work of individual researchers. AI research is also divided by several technical issues. Some subfields focus on the solution of specific problems. Others focus on one or several possible approaches or on the use of a particular tool or towards the accomplishment of particular applications.

The central problems (or goals) of AI research include reasoning, knowledge, planning, learning, natural language processing (communication), perception and the ability to move and manipulate objects.\(^7\) General intelligence is still among the field’s long-term goals.\(^8\) Currently popular approaches include statistical methods, computational intelligence and traditional symbolic AI. There are a large number of tools used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics, and many others. The AI field is interdisciplinary, in which a number of sciences and professions converge, including computer science, mathematics, psychology, linguistics, and specialized fields such as artificial psychology.

The field was founded on the claim that a central property of human intelligence—the intelligence of Homo sapiens sapiens—can be so precisely described that a machine can be made to simulate it. This raises philosophical arguments about the nature of the mind and the ethics of creating artificial beings endowed with human-like intelligence, issues which have been explored by myth, fiction and philosophy since antiquity. Artificial intelligence has been the subject of numerous optimism\(^9\) but has also suffered startling setbacks.\(^10\) Today it has become an essential part of the technology industry, providing the heavy lifting for many of the most challenging problems in computer science.\(^11\)
Representation of phenotypes
Difficulties: Rare or out-of-vocabulary words

- Google News: 100B tokens
  - 60% token coverage in HPO ontology
- PubMed: 4B
  - 91%

Human Phenotype Ontology
(Contains 11,591 phenotypes)

Bronchomegaly
Macromandible, Mandible prognathism
Absent nares, Abouphalia
Hyperthreoinuria
Milphosis, Loss of eyelashes
Butterfly vertebrae, Schmetterlingswirbel
Myodeopsia, Vitreous floaters
Ocal seizures, afebril
Micromaxilla, hypoplasia of the maxilla
Lytic defectof terminal phalanx of middle finger
Representation of phenotypes
Difficulties: Domain specific usages

Bilateral **digit** clinodactyly

1. **digit**, figure (one of the elements that collectively form a system of numeration)

2. finger, fingerbreadth, finger's breadth, **digit** (the length of breadth of a finger used as a linear measure)

3. **digit**, dactyl (a finger or toe in human beings or corresponding body part in other vertebrates)
Representation of phenotypes
Difficulties: Domain specific usages

Hypoplasia of the radius

1. radius, $r$ (the length of a line segment between the center and circumference of a circle or sphere)

2. radius (a straight line from the center to the perimeter of a circle (or from the center to the surface of a sphere))

3. radius (a circular region whose area is indicated by the length of its radius) "they located it within a radius of 2 miles"

4. radius (the outer and slightly shorter of the two bones of the human forearm)

5. spoke, wheel spoke, radius (support consisting of a radial member of a wheel joining the hub to the rim)
Representation of phenotypes

Difficulties: Compositionality

Asymmetric limb shortening
Abnormality of limb bone morphology
  ↓
Abnormality of digit
  ↓
Abnormality of finger
  ↓
Abnormality of phalanx of finger

Abnormality of the phalanges of the 5th finger
  ↓
Abnormality of the distal phalanx of the 3rd toe

Abnormality of the distal phalanx of the 5th finger
  ↓

Hypoplasia of the distal phalanx of the 5th toe
Second order representation
Rich semantic representations for domain-specific lexical

1. Disambiguate
2. Get additional context from Wikipedia
3. Compute a set of representative words
4. Construct a rich representations based on pre-trained vectors

Rare or OOV
Second order representation
Disambiguate phenotypes
Anisocoria

Mechanical anisocoria

Pupillary response

Anisocoria: All size of pupil is unequal. May be: drug induced, neurological, ophthalmic causes.

- Mechanical anisocoria: Pupil continues to react to light. Causes: local trauma, foreign body.
- Mechanical anisocoria: Pupil does not react to light. Causes: anticholinergic drugs, pontine, 3rd nerve palsy.
- Mechanical anisocoria: Pupil does not react to light. Causes: pupil toxicity, CNS depression, drug induced, sympathetic paralysis.

Conditions:
- Miosis: = Small pupil size.
- Mechanical anisocoria: = Pupil changes size.
- Miosis: = Pupil small size.
- Pupillary mechanism: = Pupil constricts.

- Mechanical anisocoria: = Pupil constricts.
- Mechanical anisocoria: = Pupil constricts.
- Pupil mechanism: = Pupil constricts.
- Mechanical anisocoria: = Pupil constricts.
- Pupillary mechanism: = Pupil constricts.
- Anisocoria: = Small pupil size.
- Miosis: = Large pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Large pupil size.
- Pupil mechanism: = Small pupil size.
- Mechanical anisocoria: = Small pupil size.
Second order representation
Lexical specificity
contrast word frequencies across relevant articles and all Wikipedia articles

\[ \text{Spec}(H; h; G; g) = -\log_{10} P(X \geq g) \]

... and extract a set of weighted words that can best represent this context
Second order representation
Representative words, example 1

Anisocoria unequal pupil size, may represent a benign physiologic variant or a manifestation of disease.
Chronic recurrent multifocal osteomyelitis a type of osteomyelitis (bone infection)
Second order representation
Combine the representations of representative words!

\[ v_i = \sum_{j=1}^{m} e^{-\lambda_j} \nu_j(r_j)_i \quad i = 1, \ldots, n \]
Second order representation

Example: the most similar phenotypes (among 11,591).

<table>
<thead>
<tr>
<th>sim.</th>
<th>Flexion contracture of digit</th>
<th>sim.</th>
<th>Bipolar affective disorder</th>
<th>sim.</th>
<th>Chaotic rapid conjugate ocular movements</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.94</td>
<td>Flexion contracture of finger</td>
<td>0.80</td>
<td>Personality disorder</td>
<td>0.85</td>
<td>Abnormal conjugate eye movement</td>
</tr>
<tr>
<td>0.92</td>
<td>Flexion contracture of thumb</td>
<td>0.85</td>
<td>Schizophrenia</td>
<td>0.80</td>
<td>Jerky ocular pursuit movements</td>
</tr>
<tr>
<td>0.91</td>
<td>Congenital finger flexion contractures</td>
<td>0.85</td>
<td>Psychosis</td>
<td>0.76</td>
<td>Slow saccadic eye movements</td>
</tr>
</tbody>
</table>

sim. **Hydranencephaly** (A defect of development of the brain characterized by replacement of greater portions of the cerebral hemispheres [...].)

<table>
<thead>
<tr>
<th>sim.</th>
<th>Porencephaly</th>
<th>sim.</th>
<th>Dandy-walker malformation</th>
<th>sim.</th>
<th>Ventriculomegaly</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81</td>
<td>(A disorder of the brain in which a cyst or cavity filled with cerebrospinal fluid develops in the cerebral hemisphere.)</td>
<td>0.79</td>
<td>(A congenital brain malformation typically characterized by incomplete formation of the cerebellar vermis, dilation of [...].)</td>
<td>0.77</td>
<td>(An increase in size of the ventricular system of the brain.)</td>
</tr>
</tbody>
</table>
Experiment
Synonym identification

Neurilemmoma : Schwannoma

List of most similar items to “Neurilemmoma”

<table>
<thead>
<tr>
<th>HP</th>
<th>Title</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP:0100008</td>
<td>schwannoma</td>
<td>1</td>
</tr>
<tr>
<td>HP:0100011</td>
<td>scleral schwannoma</td>
<td>0.8684</td>
</tr>
<tr>
<td>HP:0009593</td>
<td>peripheral schwannoma</td>
<td>0.8602</td>
</tr>
<tr>
<td>HP:0009588</td>
<td>vestibular schwannoma</td>
<td>0.8283</td>
</tr>
<tr>
<td>HP:0001067</td>
<td>neurofibromas</td>
<td>0.7845</td>
</tr>
<tr>
<td>HP:0006751</td>
<td>paraspinal neurofibromas</td>
<td>0.7713</td>
</tr>
<tr>
<td>HP:0009732</td>
<td>plexiform neurofibroma</td>
<td>0.7697</td>
</tr>
<tr>
<td>HP:0100698</td>
<td>subcutaneous neurofibromas</td>
<td>0.7659</td>
</tr>
<tr>
<td>HP:0009589</td>
<td>bilateral vestibular schwannoma</td>
<td>0.761</td>
</tr>
<tr>
<td>HP:0009590</td>
<td>unilateral vestibular schwannoma</td>
<td>0.7381</td>
</tr>
<tr>
<td>HP:0007524</td>
<td>atypical neurofibromatosis</td>
<td>0.7049</td>
</tr>
<tr>
<td>HP:0002858</td>
<td>meningioma</td>
<td>0.7048</td>
</tr>
<tr>
<td>HP:0012033</td>
<td>sacral lipoma</td>
<td>0.7039</td>
</tr>
<tr>
<td>HP:0002668</td>
<td>paraganglioma</td>
<td>0.6976</td>
</tr>
<tr>
<td>HP:0030430</td>
<td>neuroma</td>
<td>0.6945</td>
</tr>
<tr>
<td>HP:0002888</td>
<td>lipoma</td>
<td>0.6833</td>
</tr>
<tr>
<td>HP:0100697</td>
<td>neurofibrosarcoma</td>
<td>0.6812</td>
</tr>
<tr>
<td>HP:0002888</td>
<td>ependymoma</td>
<td>0.6771</td>
</tr>
<tr>
<td>HP:0001010</td>
<td>spinal meningioma</td>
<td>0.6737</td>
</tr>
<tr>
<td>HP:0001031</td>
<td>subcutaneous lipoma</td>
<td>0.6598</td>
</tr>
<tr>
<td>HP:0100635</td>
<td>carotid paraganglioma</td>
<td>0.6567</td>
</tr>
</tbody>
</table>

.... Then check for the rank of “Neurilemmoma” in the list, the higher this in the list the better!

HPO statistics:
Total phenotypes: 11,591
~7400 synonyms for phenotype titles
Experimental Results

Mean rank

First match percentage
Thank you!