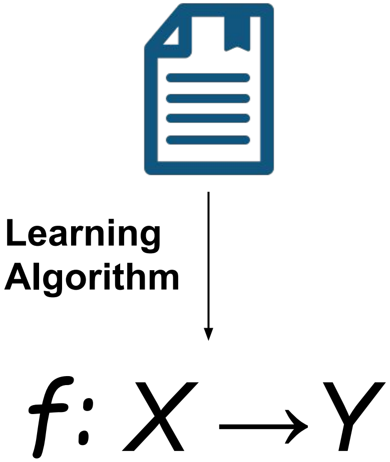
Commonsense Reasoning: Knowledge Acquisition

Never-Ending Language Learner (NELL)

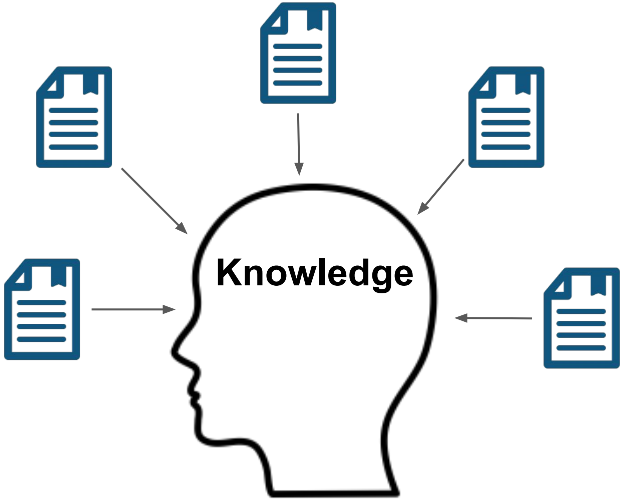
Contents

What is NELL?

Motivation of Never-Ending Learning



Machine learning



Human learning

Never-Ending Learning

- Tenet1: Natural Language Understanding requires a belief system.
  - With the belief system, a machine can react to arbitrary sentences.
- Tenet2: We will never really understand learning unless we build a machines that:
  - learn many different things
  - over years
  - and become better learners over time

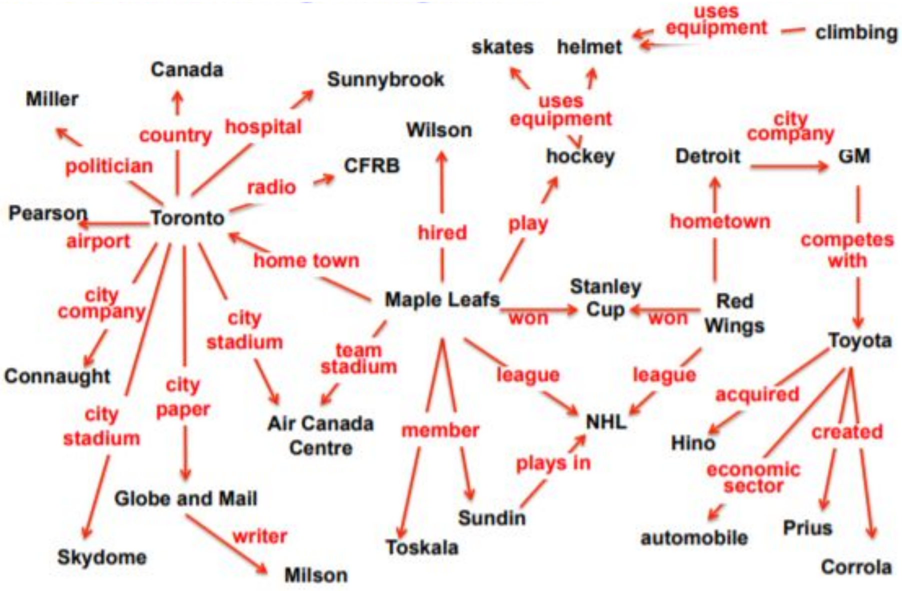
Never-Ending Learning

- "Informally, we define a never-ending learning agent to be a system that, like humans, learns **many types** of knowledge, from years of diverse and **primarily self-supervised** experience, **using previously learned knowledge to improve subsequent learning**, with sufficient self-reflection to avoid plateaus in performance as it learns."

Never-Ending Language Learner (NELL)

- NELL is a case study of Never-Ending learning.
- NELL reads the web and learns an ontology including **categories** (e.g., Sport, Athlete) and **binary relations** (e.g., AthletePlaysSport(x,y)).
- NELL is initialized with a dozen labeled training examples (e.g., Sport(baseball), Sport(soccer)) and 500M web pages (clue web), and has access to web search API and human interaction (~5mins/day). https://twitter.com/cmunell?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor
- NELL runs 24/7, forever, to **extract information** from the web to **improve its knowledge base**.
- 120M beliefs has been learned by the time the paper is written.

NELL Knowledge Fragment



Each edge represents a belief triple (e.g., play(MapleLeafs, hockey), with an associated confidence and provenance not shown here. (Figure from the paper.)

An example of NELL: ‘diabetes’

NELL believes ‘diabetes’ is a physiological condition for a number of contexts it extracts, e.g.,

- *doctor, who is diagnosed with <u>diabetes</u>*
- *preventable illnesses such as <u>diabetes</u>*
- *daughter was very sick with <u>diabetes</u>*

Each of the contexts provide a probability that diabetes is a physiological condition, together this is overwhelming evidence that.... An interesting thing is NELL is not initialized with these contexts. NELL actually learns them during these years. (so far it has ~0.5M such context patterns)

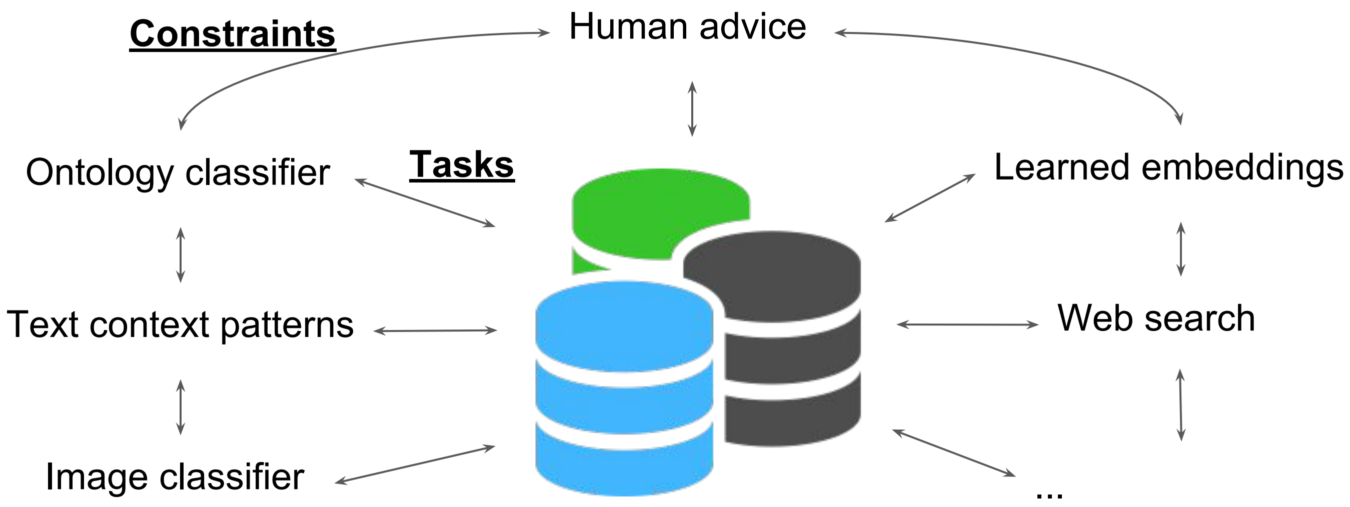
An example of NELL: ‘diabetes’

NELL usually has many beliefs about a noun phrase, e.g.,

- Mice, cats, dogs, children, people can get diabetes.
- ‘diabetes’ is a disease associated with emotion numbness.
- ‘diabetes’ can be caused by carbohydrates, glucose, junk food, and sugar levels (indicator?) (sugar?).
- Foods like vegetables can decrease the risk of ‘diabetes’.
- ‘diabetes’ can possibly be treated by the drug Avandia or glucophage.
- ...

How does NELL obtain and what does it do with its knowledge base?

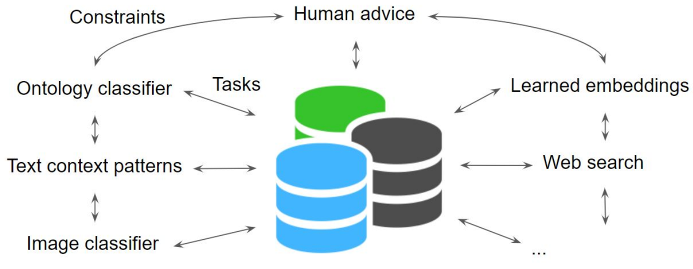
NELL's architecture



NELL Knowledge base

NELL’s tasks

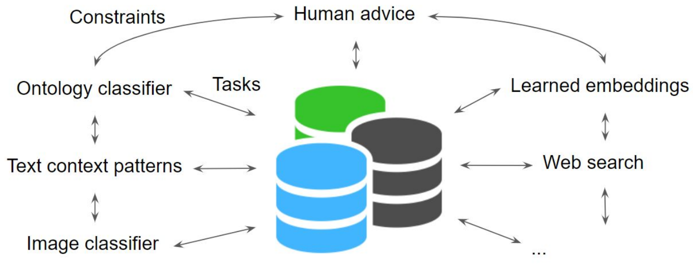
- Category classification
  - Classify noun phrase by semantic category.
- Relation classification
  - Classify noun pairs by relation.
- Entity resolution
  - Classify noun pairs as synonyms.
- ...
- In total 4100 different tasks which fall into several groups



NELL Knowledge base

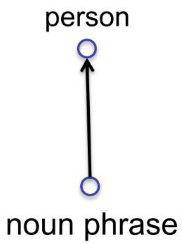
NELL's coupling constraints

- Multi-view
  - Two different views should predict the same label.
- Subset/superset
  - Categories should have immediate and super parents.
- Multi-label mutual exclusion
  - Some categories are not compatible with each other.
- Relation-argument type
  - Argument type must meet the relation requirements.
- Horn clause
  - Horn clause rule. (clause A→ B)
- In total over 1M coupling constraints, **learned by data-mining**

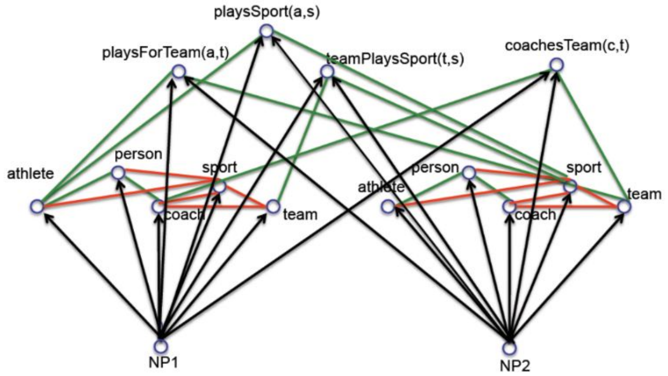


NELL Knowledge base

Example: Ontology classification, multi-task learning



hard
(underconstrained)
semi-supervised
learning problem



much easier (more constrained)
semi-supervised learning problem

Those figures are from Mitchell’s presentation.

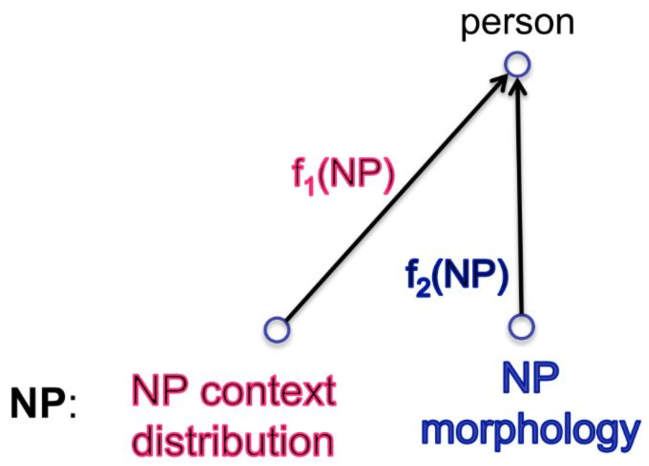
Example: Ontology classification, multi-view learning

Supervised training of one function:

Minimize: $$\sum_{np \,\in\, labeled\ data} |f_1(np) - person_{np}|$$

Supervised training of two coupling function:

Minimize: $$\sum_{np \,\in\, labeled\ data} |f_1(np) - person_{np}| + \sum_{np \,\in\, labeled\ data} |f_2(np) - person_{np}| + \sum_{np \,\in\, unlabeled\ data} |f_1(np) - f_2(np)|$$



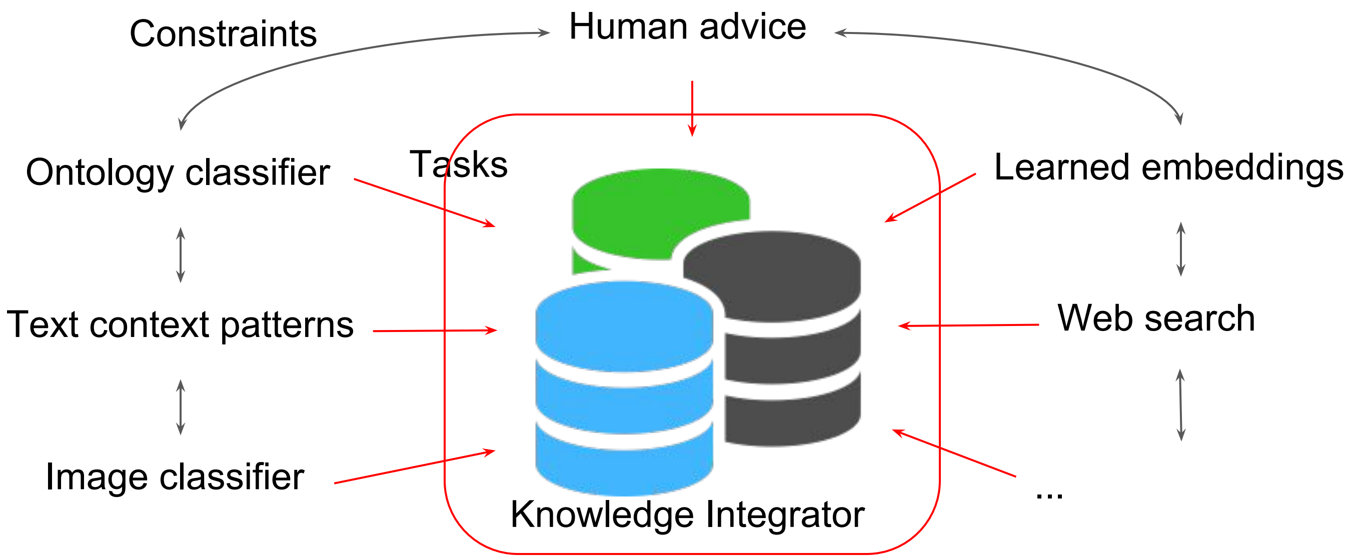
How does NELL improve (learn) its knowledge base given its architecture?

NELL's learning as Expectation Maximization (EM)

EM algorithm: Learn estimation of parameters when the model has latent variables.
Initialize parameters, then repeat until convergence:
E-step:
  Compute and update latent variables using current parameter estimation.
M-step:
  Update the parameters with MLE using current latent variable estimation.

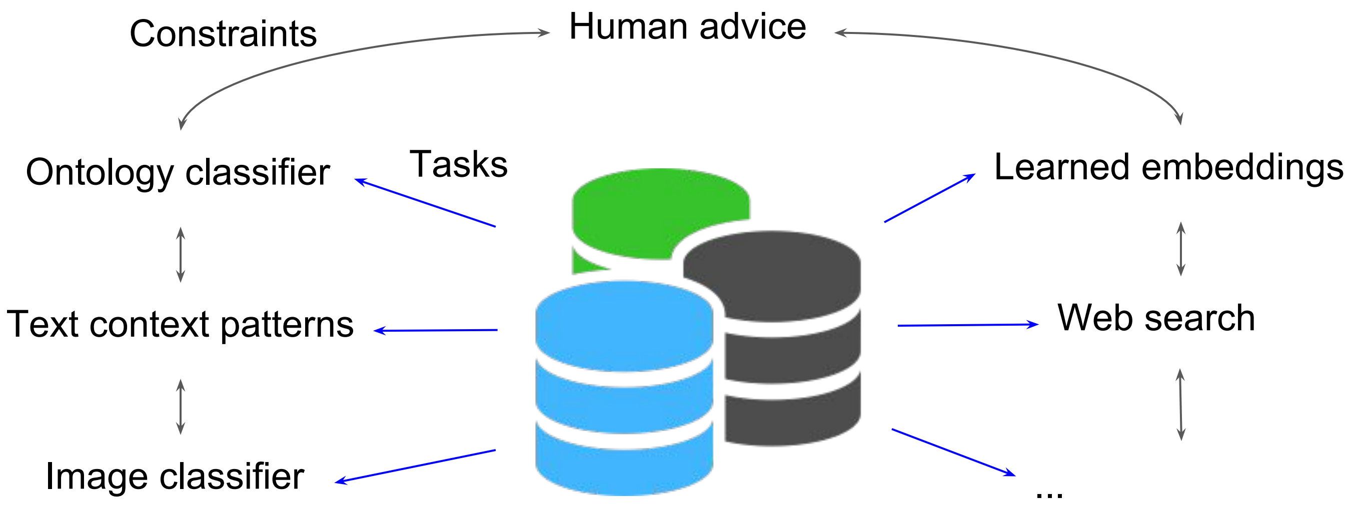
- The learning of NELL is a **semi-supervised bootstrapping learning**.
- NELL can be seen as an infinite loop of an EM algorithm.
- All the learning tasks can be seen as the parameters.
- The knowledge base can be seen as the shared latent variables.

NELL’s E-step learning



Update NELL Knowledge base

NELL's M-step learning



Retrain models using NELL Knowledge base

Thats it! (NELL's EM learning)

NELL’s Ontology Extension (OntExt)

NELL does not fix its ontology, rather it discovers new relations over time.

Approach: (Mohamed et al. EMNLP 2011)

For each pair of categories C1, C2:

cluster pairs of instances in terms of contexts that ‘connect’ them.

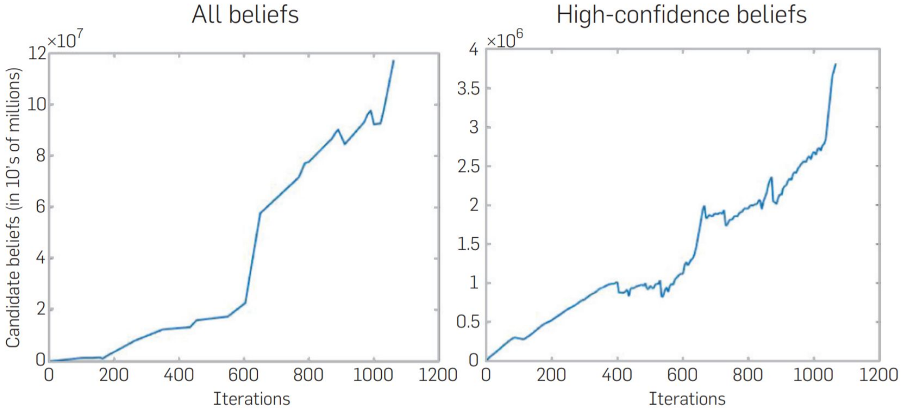
e.g., Musician and MusicInstrument has contexts:

ARG0 plays ARG1   *ARG1 master ARG0*   *ARG1 legend ARG0*

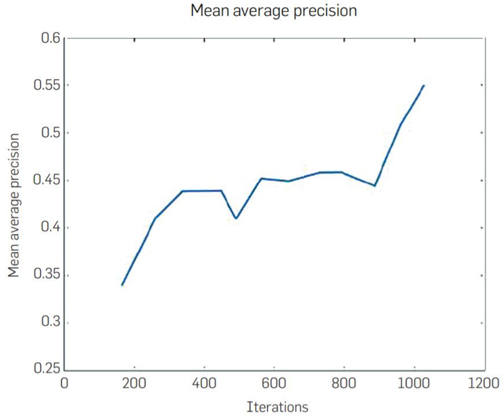
Relations generated by OntExt

- athleteWonAward
- animalEatsFood
- languageTaughtInCity
- clothingMadeFromPlant
- beverageServedWithFood
- fishServedWithFood
- athleteBeatAthlete
- plantRepresentsEmotion
- foodDecreasesRiskOfDisease

Evaluation of NELL



NELL's KB size over time. NELL's KB keeps growing over time, although only 3% of the knowledge is of high-confidence.



A test of NELL's reading accuracy by predicting novel instances of certain categories and relations.

Lessons from NELL

To better learn a never-ending learning system:

- Couple the training of many different learning tasks.
- Allow the model to learn additional coupling constraints.
  - NELL can learn new Horn clause by data-mining the KB.
- Allow the model to learn new representation beyond the initial representation.
  - NELL has the ability to suggest new relations between existing categories. (e.g., RiverFlowsThroughCity(x,y) between river, city)
- Organize learning tasks from easy to difficult.

NELL’s limitations

- NELL is not aware of how well it does.
  - NELL cannot detect that knowledge about certain areas are already saturated. E.g., Country
- Some parts of NELL are not open to learning.
  - This puts NELL under the risk of reaching a performance plateau.
- Lack of powerful reasoning components.
  - NELL currently lacks the ability to represent and reason about time and space.

NELL’s conceptual and theoretical problems

- The relationship between consistency and correctness.
    - Is an increasingly consistent learning agent also an increasingly correct agent?
    - Under what conditions is it correct?
- Convergence guarantees in principle and in practice.
    - What kind of architecture is sufficient to guarantee that the agent will converge to high performance without hitting performance plateaus.

References

Never-Ending learning

T Mitchell, W Cohen, E Hruschka, P Talukdar, J Betteridge ..., AAAI, 2015

Never-ending learning *

T Mitchell, W Cohen, E Hruschka, P Talukdar, B Yang, ..., Communications of the ACM, 2018

What Never Ending Learning (NELL) Really is? - Tom Mitchell

https://www.youtube.com/watch?v=MUMkrhrDmqQ, https://drive.google.com/file/d/0B_G-8vQI2_3QeENZbVptTmY1aDA/view

Thanks