

Coupled Semi-Supervised Learning for Information Extraction

Andrew Carlson, Justin Betteridge, Richard C. Wang,
Estevam R. Hruschka Jr., Tom M. Mitchell

Presented by : Carlos Perez

OBJECTIVE

Extract instances and predicates from Webpages.

Examples of instances for class Creature: Dragon, Gnome, Troll, Giant squid

Example of predicate

“arg1 throws fire” “arg1 and arg2 live in the forest” “arg1 lives in the pond”

Propose a methodology to circumvent semantic drift problem: Predictions get more imprecise each iteration.

OUTLINE

- OBJECTIVE
- NELL PROJECT
- APPROACH
- RELATED WORK
- CPL
- SEAL
- CSEAL
- MBL
- EXPERIMENTAL EVALUATION
- RESULTS
- CONCLUSION

NELL Project

Read the Web

Research Project at Carnegie Mellon University

[Home](#)[Project Overview](#)[Resources & Data](#)[Publications](#)[People](#)

NELL: Never-Ending Language Learning

Can computers learn to read? We think so. "Read the Web" is a research project that attempts to create a computer system that learns over time to read the web. Since January 2010, our computer system called NELL (Never-Ending Language Learner) has been running continuously, attempting to perform two tasks each day:

- First, it attempts to "read," or extract facts from text found in hundreds of millions of web pages (e.g., `playsInstrument(George_Harrison, guitar)`).



Browse the Knowledge Base!

<http://rtw.ml.cmu.edu/rtw/>

Consult Knowledge Base (Categories)

NELL Knowledge Base Browser
CMU Read the Web Project

[log in](#) | [preferences](#) | [help/instructions](#) | [feedback](#)

- transportation
- port
- museum
- skiarea
- aquarium
- monument
- stadiumorevenue
- zoo
- website
 - blog
 - politicsblog
 - url
- river
- room
 - officebuildingroom
- farm
- planet
- trail
- highway
- street
- mountainrange
- cave
- zipcode
- lake
- landscapefeatures
- item
 - bodypart
 - vein
 - artery
 - bone
 - braintissue
 - muscle
 - nerve

planet
(category)
is a celestial body orbiting a star or stellar remnant that is massive enough to be rounded by its own gravity, is not massive enough to cause thermonuclear fusion, and has cleared its neighbouring region of planetesimals
View list | [map](#) | [metadata](#)
234 instances, 1 page

instance	iteration	date learned	confidence
adraslea	120	19-jun-2010	(Seed) 100.0
ashan	432	10-oct-2011	100.0
earth	120	19-jun-2010	(Seed) 100.0
gor	601	25-jun-2012	100.0
mercury	601	25-jun-2012	(Seed) 100.0
nebulos	827	02-apr-2014	100.0
planet_n Nebulos	189	16-jan-2011	100.0
planet_uranus	177	12-dec-2010	100.0
venus	120	19-jun-2010	(Seed) 100.0
planet_mercury	722	08-apr-2013	100.0
planet_venus	176	11-dec-2010	100.0
planet_earth	172	06-dec-2010	100.0
planet_mars	177	12-dec-2010	100.0
the_earth	551	19-apr-2012	100.0
planet_pluto	800	31-dec-2013	100.0
actual_earth	178	15-dec-2010	100.0
earth_two	524	05-mar-2012	99.9
planet_neptune	177	12-dec-2010	99.8
red_planet	727	26-apr-2013	99.7
earth_b	523	02-mar-2012	99.7

<http://rtw.ml.cmu.edu/rtw/>

Consult Knowledge Base (Instances)

NELL Knowledge Base Browser
CMU Read the Web Project

[log in](#) | [preferences](#) | [help/instructions](#) | [feedback](#)

categories




relations

- everypromotedthing
- abstractthing
 - event
 - convention
 - musicfestival
 - protestevent
 - meetingeventtitle
 - conference
 - micconference
 - weatherphenomenon
 - sportsevent
 - sportsgame
 - race
 - olympics
 - grandprix
 - crimeorcharge
 - earthquakeevent
 - election
 - bombingevent
 - militaryeventtype
 - militaryconflict
 - productiaunchevent
 - filmfestival
 - roadaccidentevent
 - meetingeventtype
 - eventoutcome
- mlalgorithm
- physiologicalcondition
 - disease
 - nondisease2condition

categories

- planet**(100.0%)
 - CPL @722 (98.4%) on 06-apr-2013 ["human civilizations on _" " 's magnetic field" " 's atmosphere" "civilizations on _" "space station above _" "life on _" "animals left on _" "Heaven on _" " 's Moon" "Christ lived on _"] using earth
 - CMC @1010 (83.2%) on 04-aug-2016 [PREFIX=eart 3.10315 PREFIX=ear 2.61921 SUFFIX=rth 2.58897 SUFFIX=arth 2.53201 PREFIX=ea 2.53096 SUFFIX=th 2.49354 LASTSUFFIX=rth 1.23397 FULL_POS=NN -3.09899 WORDS -4.60748 CHARS -6.13223] using earth
 - Seed
 - SEAL @175 (100.0%) on 09-dec-2010 [1 2 3] using earth

relations

- hasvisualappearance**
 - 
 - NEIL @770 (100%) on 21-sep-2013
 - 
 - NEIL @770 (100%) on 21-sep-2013
 - 
 - NEIL @770 (100%) on 21-sep-2013

Consult Knowledge Base (Relations)

NELL Knowledge Base Browser
CMU Read the Web Project

log in | preferences | help/instructions | feedback

categoriesrelations

- relatedto
 - canbethesameas
 - athletecanbethesameasactor
 - comediancanbethesameasfemale
 - scientistcanbethesameaswriter
 - ceocanbethesameascelebrity
 - comediancanbethesameasmale
 - agriculturalproductcanbethesameas
 - celebritycanbethesameasceo
 - vegetablecanbethesameasagricultural
 - jobpositioncanbethesameaspolitical
 - femalecanbethesameascomedian
 - personafriacanbethesameaswriter
 - writercanbethesameaspersonafriac
 - politicalofficecanbethesameasjobpos
 - writercanbethesameasscientist
 - actorcanbethesameasathlete
 - malecanbethesameascomedian
 - damagedinbombing
 - victiminroadaccident
 - inverseoffislandhasskiarea
 - dateof
 - dateatwhichexistslocation
 - dateevent
 - dateofmeetingeventtitle
 - dateofsportgame
 - dateoforganizandissolution
 - releasedateexpectedfordrug
 - dateofpersondeath

animaltypehasanimal
(relation: domain [animal](#), range [animal](#))

See [metadata](#) for animaltypehasanimal
1,669 instances, 1 page

See inverse [animalistypeofanimal](#) for the definition of this relation

instance	iteration	date learned	confidence
water_birds , storks	668	14-dec-2012	(Seed) 100.0
songbirds , thrushes	406	08-sep-2011	(Seed) 100.0
fowl , turkeys	670	17-dec-2012	100.0
animals (mammal), turkeys	1019	10-oct-2016	100.0
songbirds , warblers	670	17-dec-2012	(Seed) 100.0
game_birds , waterfowl (bird)	1019	10-oct-2016	(Seed) 100.0
waterbirds , waterfowl (bird)	670	17-dec-2012	(Seed) 100.0
water_birds , waterfowl (bird)	670	17-dec-2012	(Seed) 100.0
animals (mammal), wild_birds	1019	10-oct-2016	100.0
animals (mammal), wild_turkeys	1019	10-oct-2016	100.0
invertebrates (animal), copepods	1019	10-oct-2016	(Seed) 100.0
zooplankton , copepods	670	17-dec-2012	(Seed) 100.0
shrimp , copepods	698	04-feb-2016	100.0
rays (mollusk), manta_rays (crustacean)	1019	10-oct-2016	(Seed) 100.0
invertebrates (animal), mollusks	1019	10-oct-2016	100.0
organisms , mollusks	731	07-may-2013	100.0
animals (invertebrate), mollusks	746	30-jun-2013	100.0
invertebrates (animal), shrimp	1019	10-oct-2016	100.0
tunas , albacore	670	17-dec-2012	(Seed) 100.0
pelagics , barracuda	406	08-sep-2011	(Seed) 100.0
bream , bluegill (fish)	746	30-jun-2013	(Seed) 100.0

















<http://rtw.ml.cmu.edu/rtw/>

Consult knowledge base(instance of a relation)

relations

- animalistypeofanimal
 - rays (mollusk) (100.0%)
 - CPL @673 (75.0%) on 22-dec-2012 ["arg1 and spotted eagle arg2" "arg1 and eagle arg2"] using (manta_rays, rays)
 - Seed
 - OE @830 (96.5%) on 13-apr-2014 [] using (manta_rays, rays)

Facts recently learned

instance	iteration	date learned	confidence	
<u>paulo costanzo</u> is a <u>director</u>	1024	30-oct-2016	97.7	 
<u>sabi sands game reserve</u> is a <u>zoo</u>	1027	22-nov-2016	93.9	 
<u>radio tv</u> is an <u>item often found in a bedroom</u>	1025	04-nov-2016	97.5	 
<u>n1840 1864</u> is a <u>year</u>	1024	30-oct-2016	100.0	 
<u>wyman park savings and loan association</u> is a <u>bank</u>	1024	30-oct-2016	96.1	 
<u>icelandic police</u> was <u>dameged in</u> the event <u>bus</u>	1028	28-nov-2016	100.0	 
<u>chicago</u> is the <u>home city of</u> the sports team <u>northwestern university</u>	1029	03-dec-2016	100.0	 
<u>patsi kensit</u> got <u>married in</u> <u>n1997</u>	1029	03-dec-2016	100.0	 
<u>katrina</u> is a person who <u>has residence in</u> the city <u>washington d c</u>	1028	28-nov-2016	98.4	 
<u>merck</u> has <u>acquired</u> <u>medco</u>	1024	30-oct-2016	96.9	 

APPROACH

Semi-supervised learning: Use only a few labeled examples to train the model

Why not supervised learning?

Machine learning methods have been shown to be useful to extract information from structured and unstructured text.

Training of accurate extractors is costly. (Substantial number of labeled examples for training)

What's new for this paper?

Semi-supervised learning of category and relation extractors.

Training of multiple wrapper inducers by using mutual exclusion and type checking relationships.


Coupling of the inducers and extractors by assuming they make independent errors.(Mutual exclusion ,Relation Argument Type Checking,Unstructured and semi-structured text features)

Related work

Bootstrapping methods: Start with a small number of labeled “seed” examples and iteratively grow the set of labeled examples using high-confidence labels from the current model.

Coupling the learning of category extractors

Corpus

Lemur

HomeComponentsSupportAbout

ClueWeb09How to Get ItDataset DetailsRelated DataOnline ServicesIndexing with IndriWiki & EmailFAQ

The ClueWeb09 Dataset

The ClueWeb09 dataset was created to support research on information retrieval and related human language technologies. It consists of about 1 billion web pages in ten languages that were collected in January and February 2009. The dataset is used by several tracks of the [TREC](#) conference.

Dataset Specifications

Web Pages:

- 1,040,809,705 web pages, in 10 languages
- 5 TB, compressed. (25 TB, uncompressed.)

See the [Record Counts Section](#) on the [Dataset Information and Sample Files](#) page for detailed information on the distribution of records and languages.

Web Graph:

- **Entire Dataset:**
 - Unique URLs: 4,780,950,903 (325 GB uncompressed, 105 GB compressed)
 - Total Outlinks: 7,944,351,835 (71 GB uncompressed, 24 GB compressed)
- **TREC Category B (first 50 million English pages)**
 - Unique URLs: 428,136,613 (30 GB uncompressed, 10 GB compressed)
 - Total Outlinks: 454,075,638 (3 GB uncompressed, 1 GB compressed)

The web graph for both the entire dataset and for the TREC Category B dataset (first 50 million English pages) is complete. We are in the process of retrieving the data and performing the final formatting of the web graph.

Information on how the crawl progressed is also available.

Corpus

```
2113 WARC-Type: response
2114 WARC-Target-URI: http://www.locorunning.co.nz/101.shtml
2115 WARC-Warcinfo-ID: f8ea7d54-e7a3-4d33-9ff4-45d99aa7864c
2116 WARC-Date: 2009-03-6T14:59:49-0700
2117 WARC-Record-ID: <urn:uuid:eeac9cc6-f630-410f-9a2d-8dc65f59a23d>
2118 WARC-TREC-ID: clueweb09-en0039-05-00016
2119 Content-Type: application/http;msgtype=response
2120 WARC-Identified-Payload-Type:
2121 Content-Length: 13586
2122
2123 Accept-Ranges: bytes
2124 Content-Type: text/html
2125 Server: Apache
2126 Date: Tue, 27 Jan 2009 11:10:23 GMT
2127 Connection: close
2128 Content-Length: 13431
2129
2130 <html>
2131 <head>
2132 <TITLE>LocoRunning New Zealand - Buy running shoes online</TITLE>
2133 <meta name="description" content="LocoRunning New Zealand sells running shoes designed by runners for runners of all abilities in New Zealand." >
2134 <meta name="keywords" content="mojo perfecto bandito perfecto running run shoes sell training marathon runners jogging jogger joggers runner run performance" >
2135 <meta name="htdig-keywords" content="mojo perfecto bandito perfecto running run shoes sell training marathon runners jogging jogger joggers runner run performance" >
2136 <meta name="copyright" content="Copyright &copy; 2005 LocoRunning New Zealand Ltd All rights reserved.">
2137 <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
2138 <script type="text/javascript">
2139 <!--
2140
2141 function newImage(arg) {
2142   if (document.images) {
2143     rslt = new Image();
2144     rslt.src = arg;
2145     return rslt;
2146   }
2147 }
2148
2149 function changeImages() {
2150   if (document.images && (preloadFlag == true)) {
2151     for (var i=0; i<changeImages.arguments.length; i+=2) {
```

Input Ontology

Classes

- Celestial_body
 - Star
 - Planet
- Creature
 - Mythological
 - Mythological_creature
- Group
 - Organization
- Element

Relations

- PlanetOrbitsStar(Planet,Star)
- CreatureLivesIn(Creature,Planet)
- IsRuledBy(Planet,Organization)
- IsComposedOf(Celestial_body,Element)

CPL

Predicates

p1: planet known as arg1

p2: star known as arg1

p3: planet arg1 is ruled by
arg2

Mutual Exclusion

p1 and p2 are mutually exclusive
arguments

Relation Argument Type Checking

planetIsRuledBy(Planet,Organization)

starKnownAs(Star)

planetKnownAs(Planet)

CPL

Instances

Mysterious_creature : Creature (Seed)

Sirio: Star (Seed)

Daneb:Star (Seed)

Polux :: Star (Seed)

Mercury:Planet (Seed)

Venus:Planet (Seed)

CPL

<html>

.... I saw the beautiful rings of Saturn.

There was once an isolated planet known as nebulos. ...

In the solar system there is a star known as sun but in the world of microblogging there is a star known as tumblr. ...

There is a type of planet known as Ice giants. ...

Many people think planet Earth is ruled by intelligent dinosaurs or that planet Earth is ruled by a mysterious creature but apparently it is very likely that planet Earth is ruled by a single government, nobody knows for sure.

CPL

Extract candidate instances (Categories)

- p1: planet known as arg1
 - {Ice_giants, earth, nebulos, earth, earth, nebulos, HD 20782,earth,nebulos.}
- p2: star known as arg1
 - {Sun,Nemesis, Tumblr, HD 20782}
- p3: planet arg1 is ruled by arg2
 - {<Earth,Misterious_creature>,<Nebulos, Nebulans>,<Earth,single government>,<Nebulos, Nebulans> ,<Earth, Intelligent_dinosaurs>,<Nebulos, Nebulans>}

CPL

<html>

.... I saw the beautiful rings of Saturn.

Men are from Mars and Women from Venus broke the sales record for new bestsellers.

Extract candidate instances (Predicates)

- p4?: beautiful rings of arg1
- p5?: arg1 broke the sales record
 - {Sun,Nemesis, Tumblr, HD 20782}

CPL

Filter candidates that violate coupling (Mutual exclusion)

- p1: planet known as arg1
 - {Ice_giants, earth, nebulos, earth, earth, nebulos,earth,nebulos.}- **HD 20782**
- p2: star known as arg1
 - {Sun,Nemesis, Tumblr,nebulos}- **HD 20782**
- p3: planet arg1 is ruled by arg2
 - {<Earth,Misterious_creature>,<Nebulos, Nebulans>,<Earth,single government>,<Nebulos, Nebulans> ,<Earth, Intelligent_dinosaurs>,<Nebulos, Nebulans>}

CPL

Filter candidates that violate coupling (Relation Argument Type Checking)

- p1: planet known as arg1
 - {Ice_giants, earth, nebulos, earth, earth, nebulos,earth,nebulos.}
- p2: star known as arg1
 - {Sun,Nemesis, Tumblr,nebulos}
- p3: planet arg1 is ruled by arg2
 - {<Nebulos, Nebulans>,<Earth,single government>,<Nebulos, Nebulans> ,<Earth, Intelligent_dinosaurs>,<Nebulos, Nebulans>} -<Earth,Misterious_creature>,

CPL

Filter candidates that violate coupling (Co-occurrence)

- p1: planet known as arg1
 - { earth, nebulos, earth, earth, nebulos,earth,nebulos.}-**Ice_giants**
- p2: star known as arg1
 - {Sun}-**Nemesis,Tumblr,nebulos**
- p3: planet arg1 is ruled by arg2
 - {<Nebulos, Nebulans>,<Nebulos, Nebulans> ,<Nebulos, Nebulans>}-**<Earth,single government>, <Earth, Intelligent_dinosaurs>**

CPL

Rank

- p1: planet known as arg1
 - { earth, nebulos, earth, earth, nebulos, earth, nebulos. }
- p2: star known as arg1
 - {Sun}
- p3: planet arg1 is ruled by arg2
 - {<Nebulos, Nebulans>, <Nebulos, Nebulans>, <Nebulos, Nebulans>}

Calculate Precisions

- p1
 - earth 4/9
 - nebulos 3/9
- p2
 - sun 1/4
- p3:
 - <Nebulos, Nebulans> 3/6

CPL

Rank

1. <Nebulos, Nebulans> 50%
2. earth 44.4%
3. nebulos 33.3%
4. sun 25%

CPL

Rank

1. <Nebulos, Nebulans> 50%
2. earth 44.4%
3. nebulos 33.3%
4. sun 25%

Promote at most 100 instances. In this particular example restriction will be set to at most two.

CPL

Instances

Mysterious_creature : Creature (Seed)

Sirio: Star (Seed)

Daneb:Star (Seed)

Polux : Star (Seed)

Mercury:Planet (Seed)

Venus:Planet (Seed)

Nebulus:Planet

earth:Planet

Nebulans:Organization

CPL

Predicates

p1: planet known as arg1

p2: star known as arg1

p3: planet arg1 is ruled by
arg2

Mutual Exclusion

p1 and p2 are mutually exclusive
arguments

Relation Argument Type Checking

planetIsRuledBy(Planet,Organization)

starKnownAs(Star)

planetKnownAs(Planet)

SEAL

URL: http://starwars.wikia.com/wiki/List_of_planets

```
<tr>
<td><b><a href="/wiki/Aargonar" title="Aargonar">A
</td></tr>
<tr>
<td><b><a href="/wiki/Abafar" title="Abafar">Abafa
</td></tr>
<tr>
<td><b><a href="/wiki/Abednedo" title="Abednedo">A
</td></tr>
<tr>
<td><b><a href="/wiki/Abhean/Canon" title="Abhean/
</td></tr>
<tr>
```

Seeds

Aargonar:Planet

Agamar:Planet

Solan:Planet

SEAL

Wrapper constructed by SEAL: `<td><a href="/wiki/arg1"`

```
<tr>
<td><b><a href="/wiki/Aargonar" title="Aargonar">A
</td></tr>
<tr>
<td><b><a href="/wiki/Abafar" title="Abafar">Abafa
</td></tr>
<tr>
<td><b><a href="/wiki/Abednedo" title="Abednedo">A
</td></tr>
<tr>
<td><b><a href="/wiki/Abhean/Canon" title="Abhean/
</td></tr>
<tr>
```

Content

Aargonar, Abafar,
Abednedo, Abhean/Canon,
Absanz, Affa, Agamar...

...

...Zhadalene, Solan, Zygerria

Input Ontology

Classes

- Celestial_body
 - Star
 - Planet
- Creature
 - Mythological
 - Mythological_creature
- Group
 - Organization
- Element

Relation

- PlanetOrbitsStar(Planet,Star)
- CreatureLivesIn(Creature,Planet)
- IsRuledBy(Planet,Organization)
- IsComposedOf(Celestial_body,Element)

CSEAL

URL1: some url Domain: celestial_bodies

Wrapper <td><a href="/planet/arg1"

Aargonar, Nebulus, Earth, Solan

URL2: some url Domain: celestial_bodies

Wrapper <a>
<a href="/cbody/arg1"

Mars,Polux

URL3: some url Domain: celestial_bodies

Wrapper <td><a href="/planet/arg1"

Aargonar, Nebulus, Earth,
Solan,isRuledBy(Earth,trade)

URL4: some url Domain: celestial_bodies

Wrapper <a>
<a href="/cbody/arg1"

Mars,Sun,Orbits(Mars,Sun)

CSEAL filtering

URL1: some url Domain: celestial_bodies

Wrapper <td><a href="/planet/arg1"

Aargonar, Nebulus, Earth, Solan

URL2: some url Domain: celestial_bodies

Wrapper <a>
<a href="/cbody/arg1"

Mars,Polux

URL3: some url Domain: celestial_bodies

Wrapper <td><a href="/planet/arg1"

Aargonar, Nebulus, Earth,
Solan,isRuledBy(Earth,trade)

URL4: some url Domain: celestial_bodies

Wrapper <a>
<a href="/cbody/arg1"

Mars,Sun,Orbits(Mars,Sun)

Meta Bootstrap Learner MBL

- CPL and CSEAL are the subordinate algorithms
 1. For each predicate Planet known as arg1, Star known as.....
 2. Get promoted instances from CSEAL and CPL. They both skip promotion
 3. MBL promotes candidates.

Experimental evaluation

- Input ontology contained categories and relations from two domains.
- Categories were initialized with 15 instances and 5 seed patterns
- Relations were initialized with 15 instances 5 negative instances and no seed patterns
- Corpus 200 Million web pages, 514 million sentences.

Experimental evaluation

- 10 iterations of bootstrapping for each algorithm.
- Sample 30 instances from set of promoted instances
- Submit instances to mechanical turk for labeling
- Calculate recall
- All instances presented in lowercase

Results

Predicate	Precision (%)					Promoted Instances (#)				
	CPL	UPL	CSEAL	SEAL	MBL	CPL	UPL	CSEAL	SEAL	MBL
AcademicField	70	83	90	97	100	46	903	203	1000	181
Actor	100	33	100	97	100	199	1000	1000	1000	380
Animal	80	50	90	70	97	741	1000	144	974	307
Athlete	87	17	100	87	100	132	930	276	1000	555
AwardTrophyTournament	57	7	53	7	77	86	902	146	1000	79
BoardGame	80	13	70	77	90	10	907	126	1000	31
BodyPart	77	17	97	63	93	176	922	80	1000	61
Building	33	50	30	0	93	597	1000	57	1000	14
Celebrity	100	90	100	100	97	347	1000	72	747	514
CEO	33	30	100	77	100	3	902	322	1000	30
City	97	100	97	87	97	1000	1000	368	1000	603
Clothing	97	20	43	27	97	83	973	167	1000	102
Coach	93	63	100	83	100	188	838	619	1000	242
Company	97	83	100	100	97	1000	1000	245	1000	784
Conference	93	53	97	90	100	95	990	437	928	92
Country	57	33	97	37	93	1000	1000	130	1000	207
EconomicSector	60	23	100	10	77	1000	1000	34	1000	138
Emotion	77	53	87	60	83	483	992	183	1000	211
Food	90	70	97	80	100	811	1000	89	1000	272
Furniture	100	0	57	57	90	55	963	215	1000	95
Hobby	77	33	77	50	90	357	936	77	1000	127
KitchenItem	73	3	88	13	100	11	900	8	960	2
Mammal	83	50	93	50	90	224	1000	154	1000	169
Movie	97	57	97	100	100	718	1000	566	1000	183
NewspaperCompany	90	60	60	97	100	179	1000	1000	1000	241
Politician	80	60	97	37	100	178	990	30	1000	101
Product	90	83	-	77	70	1000	1000	0	999	127
ProductType	73	63	27	63	50	712	1000	31	1000	159
Profession	73	53	-	57	93	916	973	0	1000	171
ProfessionalOrganization	93	63	100	77	87	104	943	58	1000	163
Reptile	95	3	90	27	100	19	912	149	1000	54
Room	64	0	33	7	100	25	913	12	643	3
Scientist	97	30	100	17	100	83	971	928	1000	130
Shape	77	7	7	7	85	43	985	28	733	26
Sport	77	13	63	83	73	283	1000	225	1000	284
SportsEquipment	20	10	57	23	23	58	902	52	1000	174

Results

- Mechanical turk:
 - 96 out of 100 correctly labeled
 - Labels biased toward false negatives
- CPL obtains better precision than UPL
- CSEAL obtains better precision than SEAL
- Coupling yields better accuracies

Conclusion

Coupling circumvents the problem of semantic drift associated with bootstrap learning methods.

Large scale coupled training is a strategy to significantly improve accuracy in semi supervised learning.