AFRL-RI-RS-TR-2017-184

NEVER-ENDING LEARNING FOR DEEP UNDERSTANDING OF NATURAL LANGUAGE

CARNEGIE MELLON UNIVERSITY

OCTOBER 2017

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2017-184 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /	**/ S /**
ALEKSEY PANASYUK	MICHAEL J. WESSING
Work Unit Manager	Deputy Chief, Information Intelligence
	Systems & Analysis Division
	Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
OCTOBER 2017	FINAL TECHNICAL REPORT	OCT 2012 – APR 2017

4. TITLE AND SUBTITLE

NEVER-ENDING LEARNING FOR DEEP UNDERSTANDING OF NATURAL LANGUAGE

5a. CONTRACT NUMBER

5b. GRANT NUMBER FA8750-13-2-0005

5c. PROGRAM ELEMENT NUMBER 62303E

6. AUTHOR(S)

Tom Mitchell

5d. PROJECT NUMBER EF00

5e. TASK NUMBER 82

5f. WORK UNIT NUMBER 14

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Carnegie Mellon University
Machine Learning Department
5000 Forbes Avenue
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIEA
525 Brooks Road
Rome NY 13441-4505

DARPA
675 North Randolph St
Arlington, VA 22203-2114

10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER AFRL-RI-RS-TR-2017-184

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research has explored the thesis that very significant amounts of background knowledge can lead to very substantial improvements in the accuracy of deep text analysis and understanding. To explore this thesis we have built on our earlier research on the Never Ending Language Learning (NELL) computer system, which has been running non-stop since January, 2010, learning to read the web, and automatically constructing a large knowledge base (aka knowledge graph) by extracting structured factual assertions from unstructured text on the web.

15. SUBJECT TERMS

Knowledge base, knowledge on demand

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	20	**ALEKSEY PANASYUK**

19b. TELEPHONE NUMBER (*Include area code*) **NA**

TABLE OF CONTENTS

1.0 SUMMARY

This research has explored the thesis that very significant amounts of background knowledge can lead to very substantial improvements in the accuracy of deep text analysis and understanding. To explore this thesis we have built on our earlier research on the Never Ending Language Learning (NELL) computer system, which has been running non-stop since January, 2010, learning to read the web, and automatically constructing a large knowledge base (aka knowledge graph) by extracting structured factual assertions from unstructured text on the web. Our research pursued this thesis through two primary research thrusts:

- Significantly extend our earlier macro-reader (NELL) to grow its knowledge by learning to macro-read the web. The other is to design and implement a new software system, micro-NELL which performs deep analysis of individual sentences by drawing on the background knowledge acquired by macro-NELL,
- Design and implement a new software system, micro-NELL, which performs deep analysis of individual sentences by drawing on the background knowledge acquired by macro-NELL.

(To be precise, we use the term "macro-reading" to refer to a process of extracting information by shallow analysis of text (i.e., analysis that is short of deep semantic parsing, such as considering only the local surrounding text of a noun phrase to determine its semantic type). NELL performs this kind of macro-reading of the web, depending on the fact that many different web pages can be found that express evidence for the same fact (e.g., there are many web pages that mention FoundedCompany(BillGates, Microsoft) in different text forms, and NELL combines a shallow analysis of hundreds of millions of web pages in order to draw a high statistical confidence in the structured belief (e.g., that FoundedCompany(BillGates, Microsoft)).

In contrast, we use the term “micro-reading” to refer to the process of deep semantic analysis of individual sentences, in which redundant mentions of information is rare. For example, the sentence “Gates, Microsoft’s founder, announced yesterday that he will step down next month as CEO.” can be interpreted as asserting beliefs including FoundedCompany(BillGates, Microsoft), and WorksFor(BillGates, Microsoft), though each of these beliefs is mentioned just once, without redundancy.

In this context, our research under the current grant falls into two interwoven research threads:

- Further extending the capabilities of our NELL system to “macro-read” the web through shallow analysis of redundantly mentioned beliefs founds across hundreds of millions of web pages,
- Developing a new micro-NELL “micro reader” to perform deeper analysis of individual sentences, and relying on NELL’s macro-read background knowledge to guide the deep sentence analysis performed by this micro-NELL.

Our primary research results, described in a variety of publications and in the following pages of this report, include:

Macro-reading thrust (for an overview, see [Mitchell et al., 2015]):

- NELL’s knowledge base has grown by an order of magnitude in size and in quality, from 15 million beliefs triples at the beginning of this project, to approximately 120 million today. Furthermore, NELL’s set of high confidence beliefs has grown from approximate 350,000 at the beginning of this research project, to 3,967,568 today,
- NELL’s semi-supervised machine learning methods have improved its reading competence, over a representative set of 31 categories and relations it is attempting to read, from a Mean Average Precision (MAP) of 0.30 initially, to 0.55 today, measured over the 1000 most confident beliefs it holds regarding each of these 31 predicates – e.g., over a sample of 31,000 beliefs,
- NELL’s ability to draw inferences, to form new beliefs by applying learned inference rules, and to perform efficient inference at scale over large knowledge graphs, has improved significantly with the addition of new algorithms for inference by random-walks over the knowledge graph,
- We have introduced a new reading-on-demand functionality to NELL to enable it to read in real time to answer queries if the answer is not currently in its knowledge base, and have ported this to BBN,
- We added new reading and learning components to NELL’s macroreader, including OpenEval, which actively queries the web to determine whether to believe a given assertion, and Learned Embeddings (LE), which learns vector embeddings for each NELL entity and noun phrase and learns matrix representations of NELL’s relations to infer new category and relation instances.
- We have developed new self-reflection algorithms that can be used in NELL to evaluate the accuracy of thousands of different functions NELL is learning. Importantly, these algorithms

use unlabeled data to perform the estimate, making them useful for systems such as NELL that have relatively little labeled data.

We also explored, and published papers on topics ranging from aligning ontologies of different knowledge bases, to estimating temporal scope for extracted beliefs, to assessing the truth of extracted assertions.

Micro-reading thrust: Here we have developed micro-NELL, a system that can analyze single sentences and single passages of text in greater detail than NELL. Micro-NELL is build as a collection of components that annotate the given passage with syntactic and semantic annotations, and work as a team to perform the analysis. The key components in micro-NELL are based on the follow research advances.

- We developed an algorithm for large scale training of verb-to-relation methods. Given a very large web text corpus and a knowledge base such as NELL, this algorithm produces a verb resource that maps verbs and phrases containing verbs to the relations in the KB ontology. This is now incorporated into NELL, and uses a large extension to NELL's ontology of relations which contains tens of thousands of relations,
- We developed a probabilistic generative grammar for semantic parsing which uses NELL's KB to assign higher probabilities to semantic parses that are believed by NELL's KB, or that are at least consistent with it. Here the key idea is to use background knowledge to direct the process of semantic parsing to bias it toward interpretations of text that are consistent with background knowledge.
- We developed a set of learning algorithms for acquiring CCG semantic parsers in different problem settings, including an approach that uses an open predicate vocabulary, enabling it to produce denotations for phrases such as "Republican front-runner from Texas" whose semantics cannot be represented using the NELL or Freebase ontology. Our approach directly converts a sentence's syntactic CCG parse into a logical form containing predicates derived from the words in the sentence, assigning each word a consistent semantics across sentences,
- We developed a system for prepositional phrase attachment that resolves between alternative parses (e.g., does the prepositional phrase in "Mary caught the butterfly with the spots." attach to the verb 'caught' or to the noun 'butterfly'). This system uses a variety of types of background knowledge, including NELL's KB, to achieve improvements over existing methods,
- We developed a system for joint extraction of events and entities within a document. This Bayesian approach substantially outperforms other state-of-the-art methods for event extraction,
- We explored a variety of neural network approaches, including a multi-strategy approach to frame semantic parsing that combines two distinct neural network approaches, achieving a 5.7 F1 gain over the current state of the art for full frame structure extraction. In addition, we developed KBLSTM, a novel neural model that leverages continuous representations of NELL's KB assertions to enhance the learning of recurrent neural networks for machine reading. To effectively integrate background knowledge with information from the currently processed text, our model employs an attention mechanism with a sentinel to adaptively decide whether to attend to background knowledge and which information from KBs is

useful. Experimental results show that our model achieves accuracies that surpass the previous state-of-the-art results for both entity extraction and event extraction on the widely used ACE2005 dataset.

2.0 INTRODUCTION

For many years, researchers in natural language understanding have pointed to a key bottleneck to progress: the need for substantial background knowledge to resolve the many ambiguities inherent in natural language text and speech. One problem is that it is difficult to obtain this background knowledge. A second is that we do not yet know the precise algorithms to utilize it most effectively.

In this project we conducted research on knowledge-driven deep analysis and understanding of natural language. This research built on our earlier Darpa-funded research on never-ending language learning, in which we had developed a computer program (NELL) that has been running non-stop, 24x7 since January 2010, learning to read the web. The result, at the point we began this project, was that NELL had built up a knowledge base containing over 15 million extracted beliefs, each associated with a confidence. In addition, NELL had learned millions of extraction phrases, probabilistic parameters, and inference rules, which collectively defined NELL's continually improving reading and inference methods. NELL's continuously evolving knowledge base is browsable and downloadable at http://rtw.ml.cmu.edu.

We proposed to extend NELL in several key directions, to enable it to perform deep analysis of the types of text relevant to the DEFT program. Throughout, our approach has been based on the thesis that very significant amounts of background knowledge can lead to very substantial improvements in the accuracy of deep text analysis and understanding.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

As noted in the above Summary section, our research effort involved two interwoven threads of research: (1) extending our earlier NELL macro-reading system, and (2) developing a new micro-NELL system to read individual sentences, and to utilize NELL's background knowledge to attempt to resolve syntactic and semantic parsing ambiguities in the sentence.

The key assumptions our research was based on include:

- It is possible to build a never-ending learning system that can successfully learn many different inter-related functions in a very lightly supervised setting, an improve continually for years. In our case, this system is NELL. It has continuously grown its knowledge base, and improved its reading competence through learning, beginning in January 2010, and still operating today,
- General background knowledge can be used to improve deep semantic and syntactic analysis of individual sentences. In our case, we explored the use of NELL's background knowledge to improve the sentence-level analysis performed by micro-NELL.

3.1 NELL Extensions

Our extensions to NELL as a macroreader fall into several categories, and are described in detail in multiple publications, as summarized here:

- ***New macro-reading component: OpenEval***. We incorporated a new component to macro-read the web, which was created as a result of Mehdi Samedi's Ph.D. dissertation research [Samedi13]. This component takes a candidate belief as input, uses NELL's interface to Google's search engine to find mentions on the web of text related to that candidate belief, and uses NELL's KB to train itself in real time to extract relation instances. Essentially, OpenEval evaluates the truth of queries that are stated belief triples in NELL (e.g., DrugHasSideEffect(Aspirin, GIBleeding)). OpenEval gets a small number of instances of a predicate from NELL's KB, and uses them as seed positive examples. It automatically learns, in real time, how to evaluate the truth of a new predicate instance by querying the web and processing the retrieved unstructured web pages. In [Samedi13] it is shown that OpenEval is able to respond to the queries within a limited amount of time while also achieving high F1 score. In addition, it is shown that the accuracy of responses provided by OpenEval is increased as more time is given for evaluation. OpenEval has been extensively tested and shown empirical results that illustrate the effectiveness of this approach compared to related techniques. It is now part of NELL's continuous ongoing operation.
- ***New macro-reading component: Learned Embeddings.*** We developed and incorporated into NELL a new LE (Learned Embeddings) reading and learning module [Yang17b], which learns to embed the representations of noun phrases and also NELL's semantic categories into a continuous vector space, in which the relations between noun phrases and their categories are captured. Specifically, we employ a neural network architecture to learn a vector embedding for each noun phrase and a vector embedding for each semantic category, so that the likelihood of a relation between them is optimized. We quantify the likelihood that NELL entity X is a related by relation **r** to NELL entity Y a scoring function $\mathbf{x}\,\mathbf{M_r}\,\mathbf{y}$, where **x** and **y** are learned d-dimensional vectors (embeddings) of entities X and Y respectively, and where $\mathbf{M_r}$ is a learned d x d dimensional matrix. The scalar value produced

by $\mathbf{x} \, \mathbf{M_r} \, \mathbf{y}$ is taken as the confidence that the relation holds between these two entities. In experiments we have found the learned vector embeddings and relation matrices are especially accurate for determining which category a NELL entity belongs to (i.e., for the relation r=generalization). This new LE component is now incorporated into the routine operation of NELL's continuous run.

- ***New inference methods***. In addition to learned to extract structured beliefs from text, NELL also learns to infer new beliefs from others it has read, by learning and applying hundreds of thousands of inference rules. During this research we significantly improved on NELL's ability to learn to perform such inference. In particular, we extended NELL's ability by (a) adding new corpus statistics links to NELL's symbolic knowledge graph to significantly increase the density of links and the quality of resulting inference [Gardner13], (b) generalizing the notion of a relationship match from a discrete yes/no decision to a soft decision based on vector space similarity among learned embeddings for each NELL relation [Gardner14], and (3) further broadened the patterns in rule preconditions to support probabilistic subgraph features [Gardner15b].
- ***New reading-on-demand functionality***. Originally, NELL's reading methods were invoked routinely, but independent of incoming queries to its knowledge base. In order to respond more successfully to incoming queries we added a new functionality to NELL, enabling it to respond to queries whose answer is not in the current knowledge base by performing targeted reading on demand to attempt to answer the query by real-time reading. This component was added to NELL, and made available via a JSON web interface enabling BBN and others to access this new query/reading-on-demand system.
- ***New self-reflection algorithm to evaluate NELL accuracy from unlabeled data***. One key issue for any long-term autonomous learning system, including NELL, is that it must evaluate how it is doing. Evaluating its own performance when it only has mostly unlabeled data had been an open problem. We made very significant progress here, by developing a new algorithm that uses the *agreement rate* between different NELL components, evaluated over *unlabeled data* to produce highly accurate estimates of NELL's error rates (i.e., within a few percent of the actual error rates, even though it uses only unlabeled data). This work is described in detail in [Platanios14, Platanios16, and Platanios17]. We're now planning to incorporate this method into NELL, along with a major new component to enable self-reflection and self-direction of NELL's learning effort to target places where self-improvement is needed most.
- ***Matching ontologies across knowledge bases.*** We performed new research in methods to match ontologies (the set of categories and relations used to represent knowledge) across multiple knowledge bases. The problem of aligning ontologies and database schemas across different knowledge bases and databases is fundamental to knowledge management problems. In [Wijaya13] presented a novel approach to this ontology alignment problem that employs a very large natural language text corpus as an interlingua to relate different knowledge bases (KBs). The result is a scalable and robust method (PIDGIN) that aligns relations and categories across different KBs by analyzing both (1) shared relation instances across these KBs, and (2) the verb phrases in the text instantiations of these relation instances. Experiments with PIDGIN demonstrate its superior performance when aligning ontologies across large existing KBs including NELL, Yago and Freebase.

- ***New algorithms for assigning temporal scope to beliefs***. Although NELL is somewhat successful in extracting belief triples from the web, such as PresidentOf(US, Trump), a major difficulty in temporal scoping of these acquired beliefs; that is, determining the PresidentOf(US,Trump) holds during the particular time interval 2017 to now. We explored a number of approaches to capture temporal scope. In early work performed just prior to this research contract [Talukdar12] we developed a system that employs joint inference across multiple beliefs (e.g., PresidentOf(US,Trump) is temporally coupled to VicePresidentOf(US,Pence)). In new research performed during this contract [Wijaya14] we developed a Contextual Temporal Profiles (CTP) approach which attempts to capture both the direct statements that indicate temporal scope (e.g., "Obama was president from 2009 through 2012." and also less direct statements that indicate relevant state changes (e.g., "Obama was elected in 2008."), showing that this approach improves temporal scoping compared to early methods. Furthermore, in [Wijaya15] we extend this method and apply it to Wikipedia revision histories, demonstrating in experiments that when state-changing verbs are added or deleted from an entity's Wikipedia page text, we can predict the entity's infobox updates with 88% precision and 76% recall. One compelling application of our verbs is to incorporate them as triggers in methods for updating existing KBs, which are currently mostly static.this progress, temporal scoping remains one of NELL's greatest unsolved problems. Despite these different approaches and progress, we still consider the problem of temporal scoping of extracted beliefs to be one of the most difficult problems in large scale machine reading.
- ***Truth assessment***. Whereas NELL believes assertions that are frequently mentioned on the web, it does not attempt to explicitly evaluate the trustworthiness of assertions it reads, or their sources. To address this we developed FactChecker [Nakashole14], a language-aware approach to truth-finding. FactChecker differs from prior approaches in that it does not rely on iterative peer voting, instead it leverages language to infer believability of fact candidates. In particular, FactChecker makes use of linguistic features to detect if a given source objectively states facts or is speculative and opinionated. To ensure that fact candidates mentioned in similar sources have similar believability, FactChecker augments objectivity with a co-mention score to compute the overall believability score of a fact candidate. Our experiments on various datasets show that FactChecker yields higher accuracy than existing approaches. Despite this progress, we still view truth-evaluation as an important open problem in need of further research.

In addition to the above specific efforts, we performed additional unpublished work to develop a Portuguese and a Spanish version of NELL. Although this research is still underway, and not yet published, we can see already that NELL's core approach is working fairly well in both of these languages, and can support macro-reading in any language where word tokenization is possible. We are currently also in discussions with colleagues at Tsinghua University in Beijing regarding the possibility of developing a Chinese NELL.

3.2 Micro-NELL

In addition to the above research on macro-reading in NELL, we developed an entirely new micro-reader (i.e., a system to extract semantic information from individual sentences) which relies in various ways on NELL's background knowledge to resolve syntactic and semantic ambiguities in understanding the sentence. This is our way of pursuing the thesis that large amounts of background knowledge can be used to improve the current state of the art in sentence understanding.

Overall, the microreader, micro-NELL, consists of a set of modules that annotate the given sentence in different ways, including with belief triples compatible with NELL's ontology of relations and categories. Below are descriptions of the major components we have developed:

- ***Large scale training of multi-lingual verb-to-relation extraction methods***. In [Wijaya16b] we report on our development of a scalable algorithm that produces a database of verbs and their mappings to knowledge base (KB) relations in a given knowledge base. This is useful for extracting facts from text into the KBs, and to aid alignment and integration of knowledge across different KBs and languages. More specifically, this paper presents a scalable approach to automatically construct such a verb resource using a very large web text corpus as a kind of interlingua to relate verb phrases to KB relations. Given a text corpus in any language and any KB, it can produce a mapping of that language's verb phrases to the KB relations. Experiments with the English NELL KB and ClueWeb corpus show that the learned English verb-to-relation mapping is effective for extracting relation instances from English text. When applied to a Portuguese NELL KB and a Portuguese text corpus, the same method automatically constructs a verb resource in Portuguese that is effective for extracting relation instances from Portuguese text. This research was part of Derry Wijaya's Ph.D. dissertation, and the resulting verb extraction methods have been incorporated into micro-NELL.
- ***A generative grammar for semantic parsing***. To explore new approaches to incorporating background knowledge into micro-reading, we have developed a novel generative grammar for semantic parses that can generate sentences probabilistically, with higher probability assigned to sentences that arise from beliefs in background knowledge such as NELL's. In [Saparov17] we describe a generative process in which a logical form is sampled from a prior, and conditioned on this logical form, a grammar probabilistically generates the output sentence. Grammar induction using MCMC is applied to learn the grammar given a set of labeled sentences with their corresponding logical forms. Our semantic parser finds the logical form with the highest posterior probability exactly. We obtain strong experimental results on the standard GeoQuery dataset and achieve state-of-the-art F1 on the Jobs dataset. This component has been incorporated into micro-NELL, and thus uses NELL's background knowledge to bias its semantic parsing of new sentences.
- ***Learning CCG grammars and semantic parsers.*** One of the primary approaches to semantic parsing is the Combinatory Categorial Grammar (CCG) paradigm proposed initially by Mark Steedman. We developed this work in a number of steps [Krinamurthy13, Krishnamurthy13b, Krishnamurthy14] that formed the Ph.D. dissertation of Jayant Krishnamurthy. This culminated in [Krishnamurthy15] which presented an approach to learning a model theoretic semantics for natural language tied to Freebase. Crucially, our approach uses an open predicate vocabulary, enabling it to produce denotations for phrases such as "Republican front-runner from Texas" whose semantics cannot be represented using the Freebase schema. Our approach directly converts a sentence's syntactic CCG parse into a logical form containing predicates derived from the words in the sentence, assigning each word a consistent semantics across sentences. This logical form is evaluated against a learned probabilistic database that defines a distribution over denotations for each textual predicate.

A training phase produces this probabilistic database using a corpus of entitylinked text and probabilistic matrix factorization with a novel ranking objective function. We evaluated our approach on a compositional question answering task where it outperformed several competitive baselines. We also compared our approach against manually annotated Freebase queries, finding that our open predicate vocabulary enables us to answer many questions that Freebase cannot. This work has also been incorporated into micro-NELL.

- ***Knowledge driven prepositional phrase attachment, and information extraction from compound nouns***. Prepositional phrases (PPs) express important relational information. For example, consider the prepositional phrase information in “Mary caught the butterfly *with the spots.*” versus “Mary caught the butterfly *with the net.*” However, PPs are a major source of syntactic ambiguity and still pose problems in parsing. In [Nakashole15] we presented a method for resolving ambiguities arising from PPs, making extensive use of semantic knowledge from various resources. To train our prepositional phrase attachment algorithm to use this background knowledge, we use both labeled and unlabeled data, utilizing an expectation maximization algorithm for parameter estimation. Experiments show that our method yields improvements over existing methods including a state of the art dependency parser. This algorithm for prepositional phrase attachment is now in micro-NELL, along with a related approach to knowledge-driven extraction of relations from compound nouns. This compound noun information extractor uses background knowledge about NELL’s semantic types for various nouns to extract relations. For example, it has learned that a sequence of noun types <location><politicalOffice><person> (e.g., ‘Pittsburgh mayor Peduto’ indicates the relationship HoldsOffice(person,politicalOffice)), relying on NELL’s diverse knowledge about fine-grained semantic classes.
- ***Joint extraction of events and role fillers***. In [Yang16] we consider joint extraction of events and their related entities across the many sentences that form a document. Entities are often actors or participants in events and events without entities are uncommon. However, existing work in information extraction often models events separately from entities, and performs inference at the sentence level, ignoring the rest of the document. In [Yang16], we propose a novel Bayesian approach that models the dependencies among variables of events, entities, and their relations, and performs joint inference of these variables across a document. The goal is to enable access to document-level contextual information and facilitate context-aware predictions. We demonstrate that our approach substantially outperforms the state-of-the-art methods for event extraction as well as a strong baseline for entity extraction.
- ***Deep network approaches to semantic analysis and information extraction***. In addition to the above approaches, we have also explore deep neural network approaches to semantic analysis, including analysis based on background knowledge from NELL. In [Yang17] we introduce a new multi-strategy method for frame semantic parsing that significantly improves the prior state of the art. Our model leverages the advantages of a deep bidirectional LSTM neural network which predicts semantic role labels word by word and a relational neural network which predicts semantic roles for individual text expressions in relation to a predicate. The two networks are integrated into a single model via knowledge distillation, and a unified graphical model is employed to jointly decode frames and semantic roles during inference. Experiments on the standard FrameNet data show that our model significantly outperforms existing neural and non-neural approaches, achieving a 5.7 F1 gain over the current state of the art, for full frame structure extraction. In addition, in [Yang17b]

we consider how to take advantage of external knowledge bases (KBs) such as NELL's, to improve recurrent neural networks for machine reading. Traditional methods that exploit knowledge from KBs encode knowledge as discrete indicator features. Not only do these features generalize poorly, but they require task-specific feature engineering to achieve good performance. We developed and presented KBLSTM, a novel neural model that leverages continuous representations of KBs to enhance the learning of recurrent neural networks for machine reading. To effectively integrate background knowledge with information from the currently processed text, our model employs an attention mechanism with a sentinel to adaptively decide whether to attend to background knowledge and which information from KBs is useful. Experimental results show that our model achieves accuracies that surpass the previous state-of-the-art results for both entity extraction and event extraction on the widely used ACE2005 dataset.

4.0 RESULTS AND DISCUSSION

In our thrust on macro-reading and extensions to NELL, the most successful directions forward were (1) the addition of a reading-on-demand component which enables NELL to improve its response rate to incoming queries by reading on demand in cases where the query answer is not available by lookup in NELL's knowledge base, (2) the incorporation of Samadi's OpenEval system as an additional reading/learning component in NELL, (3) improvements in NELL's ability to infer new beliefs from old (separate from its reading components), and (4) development of novel algorithms that enable NELL and other self-learning systems to self-reflect by evaluating their accuracy based on internal consistency in how they process unlabeled data.

One of the more interesting themes to emerge in this work is the utility of learned vector embeddings for NELL. We found such vector embeddings to be useful both in NELL's new Learned Embeddings module for deciding which noun phrases refer to which semantic categories, and also in NELL's inference over its knowledge graph where learned embeddings of the knowledge graph edges/relations allow a soft match when applying inference methods. This is consistent with the more broad movement in text analysis toward greater use of such learned embeddings for words, phrases and sentences.

In addition, we made progress on problems such as extending NELL macro-reading to languages including Spanish and Portuguese, temporal scoping of beliefs, and aligning knowledge bases by using large-scale corpus data as a kind of shared grounding across the knowledge bases. Going forward, we see two of the most difficult, yet important problems are temporal scoping of extracted assertions, and determining which assertions on web text are actually factually correct.

In our thrust on micro-reading and the development of micro-NELL we explored a diverse variety of approaches, but each of these approaches was specifically chosen to explore ways in which background knowledge like NELL's can be used to improve the semantic analysis of single sentences and/or documents. Although much remains to be done, we feel our results to date already provide strong support for our underlying thesis that true understanding of text requires diverse background knowledge. We have considered approaches from CCG semantic parsing, to joint information extraction of events and associated entities, to a novel probabilistic generative grammar that used a background knowledge base such as NELL's to determine which sentence interpretations are most probable (i.e., those that are consistent with the background knowledge). We were purposely eclectic in exploring approaches, considering Bayesian approaches, traditional classifiers, and deep neural networks that incorporate learned word and sentence embeddings, embeddings of knowledge base beliefs, memory components such as LSTMs, and learned attention mechanisms. Given our evidence to date, we feel the deep network approaches are particularly attractive both because of their empirical success, and also because they offer an opportunity to integrate many processing steps into an end-to-end architecture that can be jointly learned. However, much research is now needed to study the questions of (1) which such architecture can best lead to strong language understanding, and (2) how can background knowledge acquired separately from a variety of sources (e.g., NELL, DBpedia, YAGO) best be integrated into such deep network architectures.

5.0 CONCLUSIONS

As the discussion in the above Section 4 indicates, we have made good progress in developing redundancy-based machine learning and macro-reading methods for large scale knowledge base development, and have also made significant initial strides in demonstrating that broad background knowledge is valuable in computer understanding of natural language text.

6.0 RECOMMENDATIONS

Going forward, we plan to continue our exploration of the thesis that broad-scale background knowledge can be acquired by NELL-like systems, and that this kind of background knowledge can improve the current state-of-the-art in natural language understanding. We found out very recently that we will be funded by a new DARPA grant to explore the feasibility of integrating a variety of research efforts to build large common-sense knowledge and reasoning systems, including NELL, NEIL, Yago and physic commonsense (e.g., a system for naïve physics reasoning from Josh Tennenbaum at MIT). This will help us in this direction.

Our recommendation to the Air Force and to DoD more generally is that there is a great opportunity for additional research into never-ending learning systems. There is surprisingly little research in this direction (i.e., NELL and NEIL), despite the growing need for continuous learning in embedded computer systems in many parts of the military and elsewhere.

We also recommend greater research directly targeted at discovering paradigms and algorithms by which background knowledge can provide genuine language *understanding* as opposed to shallow language processing. The practical uses of shallow language processing (e.g., for sentiment analysis, named entity extraction) are important, but given that commercial organizations are now developing many products that provide such practical types of NL *Processing*, the big opportunity for DoD is to support research on the much more ambitious goal of true NL *Understanding.*

7.0 REFERENCES

[Platanios17] Estimating Accuracy from Unlabeled Data: A Logic Approach. A. Platanios, H. Poon, E. Horvitz, and T. Mitchell, *NIPS* 2017 to appear.

[Saparov17] A Probabilistic Generative Grammar for Semantic Parsing. A. Saparov, V. Saraswat, and T. Mitchell, *CoNNL* 2017.

[Yang17] A Joint Sequential and Relational Model for Frame-Semantic Parsing. B. Yang and T. Mitchell, *EMNLP* 2017.

[Yang17b] Leveraging knowledge bases in LSTMs for improving machine reading. B. Yang, and T. Mitchell, *ACL* 2017.

[Platanios16] Estimating Accuracy from Unlabeled Data: A Bayesian Approach. Platanios, E. A.; Dubey, A.; and Mitchell, T. M. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2016.

[Yang16] Joint Extraction of Events and Entities within a Document Context
B. Yang and T. Mitchell. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2016.

[Wijaya16] VerbKB: A Knowledge Base of Verbs for Natural Language Understanding.
D. T. Wijaya, Ph.D. dissertation, Carnegie Mellon University, August 2016.

[Wijaya16b] Mapping Verbs in Different Languages to Knowledge Base Relations using Web Text as Interlingua
D. T. Wijaya and T. Mitchell. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2016.

[Gardner15] Translation Invariant Word Embeddings
M. Gardner, K. Huang, E. Papalexakis, X. Fu, P. Talukdar, C. Faloutsos, N. Sidiropoulos, T. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015 [PDF].

[Samadi15] AskWorld: Budget-Sensitive Query Evaluation for Knowledge-on-Demand
M. Samadi, P. Talukdar, M. Veloso, T. Mitchell. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015. [PDF].

[Wijaya15] "A Spousal Relation Begins with a Deletion of engage and Ends with an Addition of divorce": Learning State Changing Verbs from Wikipedia Revision History
D. T. Wijaya, N. Nakashole, T. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015.

[Gardner15b] Efficient and Expressive Knowledge Base Completion Using Subgraph Feature Extraction.
M. Gardner, T. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015.

[Nakashole15] A Knowledge-Intensive Model for Prepositional Phrase Attachment.
N. Nakashole, T. Mitchell. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. [PDF]

[Krishnamurthy15] Learning a Compositional Semantics for Freebase with an Open Predicate Vocabulary.
Jayant Krishnamurthy and Tom M Mitchell. In *Transactions of the Association for Computational Linguistics, Volume 3*, 2015. [PDF]

[Mitchell15] Never-Ending Learning.
T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, J. Welling. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, 2015. [PDF]

[Krishnamurthy14] Joint Syntactic and Semantic Parsing with Combinatory Categorial Grammar.
J. Krishnamurthy, T. Mitchell. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014.

[Betterridge14] Assuming Facts Are Expressed More Than Once.
J. Betteridge, A. Ritter and T. Mitchell. In *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference (FLAIRS-27)*, 2014.

[Platanios14] Estimating Accuracy from Unlabeled Data.
E. A. Platanios, A. Blum, T. Mitchell. In *Uncertainty in Artificial Intelligence (UAI)*, 2014.

[Wijaya14] CTPs: Contextual Temporal Profiles for Time Scoping Facts via Entity State Change Detection.
D.T. Wijaya, N. Nakashole and T.M. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[Gardner14] Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases.
M. Gardner, P. Talukdar, J. Krishnamurthy and T.M. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. [PDF]

[Nakashole14] Language-Aware Truth Assessment of Fact Candidates.
N. Nakashole, T. Mitchell. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014. [PDF]

[Gardner13] Improving Learning and Inference in a Large Knowledge-base using Latent Syntactic Cues.
Matt Gardner, Partha Pratim Talukdar, Bryan Kisiel, and Tom Mitchell. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, 2013. [PDF]

[Wijaya13] PIDGIN: Ontology Alignment using Web Text as Interlingua.
D.T. Wijaya, P.P. Talukdar and T.M. Mitchell. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, 2013. [PDF]

[Krishnamurthy13] Jointly Learning to Parse and Perceive: Connecting Natural Language to the Physical World.
Jayant Krishnamurthy and Thomas Kollar. In *Transactions of the Association for Computational Linguistics, Volume 1*, 2013. [PDF]

[Krishnamurthy13b] Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models.
Jayant Krishnamurthy and Tom M. Mitchell. In *Proceedings of the ACL 2013 Workshop on Continuous Vector Space Models and their Compositionality*, 2013.

[Samadi13] OpenEval: Web Information Query Evaluation, M. Samadi, M. Veloso, M. Blum, *AAAI*, 2013.

[Talukdar12] Acquiring Temporal Constraints between Relations.
P.P. Talukdar, D.T. Wijaya and T.M. Mitchell. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, 2012. [PDF]

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

ACE2005	Automatic Content Extraction
CCG	Combinatory Categorical Grammars
DEFT	Deep Exploration and Filtering of Text
DoD	Department of Defense
KB	Knowledge Bases
KBLSTM	Knowledge Base Long Short-Term Memory
LE	Learned Embeddings
LSTM	Long Short-Term Memory
MAP	Mean Average Precision
NELL	Never Ending Language Learning
NEIL	Never Ending Image Learning
NL	Natural Language
PP	Preposition Phrases
YAGO	Yet Another Great Ontology (open source knowledge base)