Harnessing Cognitive Information for Natural Language Processing: *An Investigation based on Eye-Tracking*

A synopsis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

by

Abhijit Mishra
(Roll No. 114056002)

Under the guidance of
Prof. Pushpak Bhattacharyya

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
2017

Abstract

Existing Natural Language Processing (NLP) systems are weak AI systems: they try to capture the *functionality* of human language processing faculty, without worrying about how such faculty is realized in the *hardware* of the human. These systems are are still agnostic of the actual cognitive processes involved in human language processing. This, however, does not deliver when applications demand order of magnitude facelift in accuracy, as well as insight into characteristics of human language processing/understanding.

In this thesis, we present our research on empowering NLP systems with cognitive modality of language processing. We employ the *eye-tracking* technology to record and analyze *shallow cognitive information* in the form of gaze patterns of readers/annotators while they perform language processing tasks. The insights gained through our analyses are translated into systems that help us to- (1) *assess the* ***cognitive effort*** *in text-annotation better, with a view to increasing human text-annotation efficiency*, and (2) *extract* ***cognitive features****, with a view to improving the accuracy of text classifiers*. Our experiments pertaining to these 2-fold objectives are documented in two different parts of the thesis, as mentioned below:

Part 1: Assessing Cognitive Effort in Text-annotation

Towards assessing the cognitive effort in human annotation, we address problems like *estimating complexity of sentence translation* and *measuring sentiment annotation complexity of short-text*. Our solutions could be useful towards (a) devising better cost models for translation/sentiment annotation crowdsourcing/outsourcing (b) choosing the right ensemble of NLP systems for machine translation/sentiment analysis. For both of these problems, we consider annotators' eye-movement information to label data with scores indicating translation complexity and sentiment annotation complexity. We rely on simple eye movement parameters like duration of *fixations* and *saccades* to measure these complexity scores. The labeled data is used to train supervised statistical regressors that predict *Translation Complexity Index (TCI)* and

Sentiment Annotation Complexity (SAC) from linguistic features extracted from the input text.

We then propose a cognitive model for measuring reading/annotation effort. The model relies on *scanpath*, a representation of eye-movement patterns that encompasses more information about both fixations and saccades than just their durations. We call the measure *Scanpath Complexity*. Scanpath Complexity is modeled as a function of various properties of gaze fixations and saccades- the basic parameters of eye movement behavior. We demonstrate the effectiveness of our scanpath complexity measure by showing that its correlation with different measures of lexical and syntactic complexity as well as standard readability metrics is better than baseline measures based on duration of fixations. Our research in this direction will have a positive impact on the text annotation process, an integral part of data-driven NLP.

While the above mentioned systems are, to some extent, effective in modeling various forms of nuances in text annotation, we observed that cognitive information captured through eye-movement patterns could also be useful to model the ability of a reader to understand/comprehend the given reading material. Specifically, during our sentiment annotation experiments, we observed a considerable variation in annotators' eye-movement behavior when *textual sarcasm* present in the text-to-be-annotated is understood *vis-à-vis* when it is not. This lead us to digress from the central goal and work on a highly specific yet important problem of *sarcasm understandability* prediction: predicting whether a reader/annotator has realized the intended meaning of a given sarcastic text or not. Sarcasm understandability is modeled as a function of various attributes extracted from the given text and eye-movement patterns of the annotators. We believe, this work lays the foundations for an even more significant problem of *personalizing text comprehensibility*.

Part 2: Extracting Cognitive Features for Text Classification

We consider eye movement patterns of annotators to derive *cognitive features*, which are used in supervised text classification tasks *viz.*, sentiment polarity and sarcasm detection along with traditional text-based features. For these two tasks, we propose two feature engineering schemes such as- (1) *handcrafting features from the text and gaze inputs by exploiting their underlying structures* and (2) *automatically learning feature-representations from the text and gaze data with the help of deep Convolutional Neural Network*. Our classifiers built on top of ensembles of text and gaze based features extracted through these schemes consistently outperform the state-of-the-art sentiment and sarcasm classifiers. We observe that incorporation of cognitive elements into traditional classifiers can better tackle linguistic subtleties (like irony, sarcasm, thwarting *etc.*), which accounts for the performance improvement. The outcome of this research will benefit web-content creators, review writers and social media analysts alike. .

Publications from the Thesis

Peer-reviewed Conferences

1. **Mishra, Abhijit** and Kanojia, Diptesh and Nagar, Seema and Dey, Kuntal and Bhattacharyya, Pushpak. 2017 Scanpath Complexity: Modeling Reading Effort using Gaze Information, **AAAI 2017** San Francisco, USA. 4-9 February, 2017.

2. **Mishra, Abhijit** and Kanojia, Diptesh and Nagar, Seema and Dey, Kuntal and Bhattacharyya, Pushpak. 2016. Harnessing Cognitive Features for Sarcasm Detection, **ACL 2016**, Berlin, Germany. 7-12 August, 2016.

3. **Mishra, Abhijit** and Kanojia, Diptesh and Nagar, Seema and Dey, Kuntal and Bhattacharyya, Pushpak. 2016. Leveraging Cognitive Features for Sentiment Analysis, **CoNLL 2016**, Berlin, Germany. 11-12 August, 2016.

4. **Mishra, Abhijit** and Kanojia, Diptesh and Bhattacharyya, Pushpak. 2016. Predicting Readers' Sarcasm Understandability by Modelling Gaze Behaviour, **AAAI 2016**, Phoenix, USA. 12-17 February, 2016.

5. Cheri, Joe and **Mishra, Abhijit** and Bhattacharyya, Pushpak. 2016. Leveraging Annotators' Gaze Behaviour for Coreference Resolution, **CogACLL 2016**, collocated with ACL, Berlin, Germany. 11 August, 2016.

6. Joshi, Aditya and **Mishra, Abhijit** and S., Nivvedan and Bhattacharyya, Pushpak. 2014. Measuring Sentiment Annotation Complexity of Text. **ACL 2014**, Baltimore, USA. 23-25 June, 2014.

7. **Mishra, Abhijit** and Joshi, Aditya and Bhattacharyya, Pushpak. 2014. A cognitive study of subjectivity extraction in sentiment annotation. **WASSA 2014**, collocated with ACL, Baltimore, USA. 27 June, 2014.

8. **Mishra, Abhijit** and Bhattacharyya, Pushpak and Carl, Michael. 2013. Automatically Predicting Sentence Translation Difficulty. **ACL 2013**, Sofia, Bulgaria. 4-9 August, 2013.

9. **Mishra, Abhijit** and Carl, Michael and Bhattacharyya, Pushpak. 2012. A heuristic-based approach for systematic error correction of gaze data for reading.*First Workshop on Eye-tracking and Natural Language Processing* (**ETNLP 2012**), collocated with COLING, Mumbai, India. 15 December, 2012.

Peer-reviewed Journals

1. **Mishra, Abhijit** and Bhattacharyya, Pushpak. 2017. Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification using Convolutional Neural Networks *Under Review- To be notified.*

Additional collaborative work

1. Joshi, Aditya and **Mishra, Abhijit** and Balamurali, AR and Bhattacharyya, Pushpak and Carman, Mark J. 2015. A computational approach for automatic prediction of drunk-texting, **ACL 2015**, Beijing, China, July 26-31, 2015.

2. Kunchukuttan, Anoop and **Mishra, Abhijit** and Chatterjee, Rajen and Shah, Ritesh and Bhattacharyya, Pushpak. 2014. Shata-Anuvadak: Tackling Multiway Translation of Indian Languages, **LREC 2014**, Rekjyavik, Iceland. 26-31 May, 2014

3. Kunchukuttan, Anoop and Chatterjee, Rajen and Roy, Shourya and **Mishra, Abhijit** and Bhattacharyya, Pushpak. 2013. TransDoop: A Map-Reduce based Crowdsourced Translation for Complex Domain (demo paper), **ACL 2013**, Sofia, Bulgaria. 4-9 August, 2013

Synopsis

Natural language processing (NLP) is concerned with the interactions between computers and human (natural) languages. Several layers of NLP are Morphological Analysis, Part of Speech Tagging, Parsing, Semantics, Discourse Analysis, and Pragmatics. NLP is applied extensively to Semantic Web, Machine Translation, Automatic Question Answering, and Summarization, *etc*.

The processing at each layer of NLP is carried out by *rule based* or *statistical (data-driven)* systems. While, in rule-based systems, the algorithms rely on a set of handcrafted rules that cover all the linguistic aspects of the tasks, statistical systems try to learn the rules as patterns from a large number of examples through mathematical modeling. Rule-based systems require a tremendous amount of manual effort which makes them cost and time unfriendly. For instance, it takes about 20 years to build a rule-based automatic translation system (like SYSTRAN) for a pair of languages. On the other hand, efficient statistical systems can be constructed quickly with less manual effort. It is due to this reason that most of the state of the art NLP systems (like Google Translator, IBMs Watson Question Answering System) are statistical in nature.

In the last decade, many statistical systems have been built, reinforced and publicized for various NLP tasks. The research trend for the last two decades points to the fact that, while it is easy to improve statistical systems from a "poor" to an "average" level, it becomes much harder to raise the accuracy further. For example, let us consider Word Sense Disambiguation, the NLP task to automatically decide the senses exhibited by the polysemous words given the context. A survey by Navigli (2009) shows that one of the earlier systems, proposed by Agirre and Rigau (1996) works with an accuracy of around 65% on general domain data for disambiguation of nouns, verbs, adverbs and adjectives. The system achieves around 13% of accuracy improvement over the pre-existing Most Frequent Sense Algorithm (52% accurate). At present, even after 18 years since Agirre and Rigau (1996) was published and, despite using knowledge-rich resources, current WSD systems accuracies have not gone beyond 72%. This hints at a possible *saturation barrier* that WSD technology may be moving towards. Similar trends have been observed for other NLP tasks like Translation and Sentiment Analysis (Liu and Zhang, 2012). This research trend has motivated NLP researchers to explore and exploit additional modalities like speech, image, stylistic patterns (such as hashtags and smilies in social media context) along with text for improving the performance of NLP systems. Modern NLP systems, we believe, are still agnostic of the actual cognitive processes involved in human language processing, and hence, can be classified as *weak AI*[1] systems.

In this thesis, we present our research on empowering NLP systems with the cognitive modality of language processing. In order to capture cognitive information that can provide insights into human language processing and understanding, we employ *eye-tracking* technology to record eye-gaze activities of human subjects during linguistic annotation. The potential of eye-tracking was realized by us after a cognitive study of word sense annotation was successfully carried out in our lab at the Center for Indian Language Technology (CFILT) by Joshi et al. (2013). We continued to pursue this research direction and leveraged eye-tracking technology to build systems pertaining to various components of NLP *viz,* annotation, classification, and evaluation. Through this thesis, we demonstrate that cognitive information collected through eye-tracking can be useful in devising better NLP systems. Our research, thus, contributes significantly towards the emergence of a new branch of NLP *i.e.*, termed as *Cognitive NLP*.

Cognitive Data: A Valuable By-product of Annotation

The basic requirement of supervised statistical methods for various NLP tasks like Part of Speech Tagging, Dependency Parsing, Machine Translation is large scale annotated data. Since statistical methods have taken places over rule/heuristic based methods over the years, there is a consistent demand of *annotation*.

Annotation[2] refers to the task of labeling of text, image or other data with comments, explanation, tags or markups. A typical annotation process involves employing professionals/linguists to label raw textual data. For each unit of data, annotation involves the following activities:

- Visualization: Viewing the raw data on a monitor screen

[1] https://en.wikipedia.org/wiki/Weak_AI
[2] http://en.wikipedia.org/wiki/Annotation

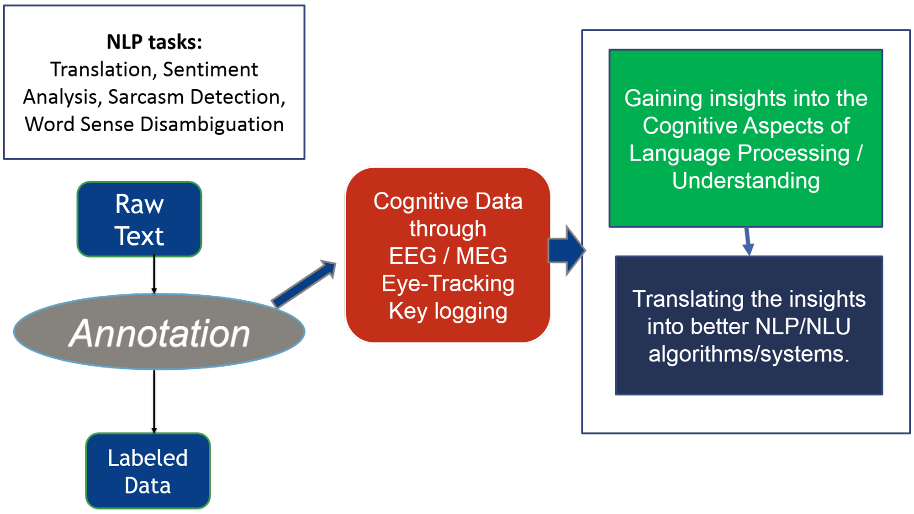

Figure 1: Cognitive Natural Language Processing: Central Idea

- Comprehension: Understanding the data through hypothesizing
- Generation: Assigning the data with appropriate information.

In most of the cases, annotators switch back and forth among these activities and finally stop at the most appropriate decision. For example, if we consider the task of translation as the process of annotating a text with its target language equivalent, we may observe that while translating, we switch between reading source and producing target segments. While the outcome of the annotation process (*i.e.,* the labeled data) matters the most, capturing the user activities during annotation is viable through various technologies available. We refer to the data representing the user activities during language processing and understanding as *Cognitive Data*. Our belief is that capturing and analyzing cognitive data may help us to: (a) *gain insights into the cognitive process underlying language processing and understanding* and (b) *translate the insights into better systems for natural language processing/understanding*. Figures 1 presents an insight into how and why cognitive information can be seamlessly harnessed in a statistical NLP setting.

How can cognitive data be captured? In most of the linguistic annotation tasks, the visualization, comprehension and generation activities are carried out through reading and writing. To capture the reading activities, one can record the eye-movement patterns through *eye-tracking technology*. Similarly, writing activities can be observed through key-logging (recording keystroke sequences). Apart from these, cognitive information can also be captured through neuro-electromagnetic signals obtained through *Electroencephalography* (EEG) (Antonenko et al., 2010), *Magnetoencephalography* (MEG) or brain-imaging obtained through *functional*

Magnetic Resonance Imaging (fMRI)(Paas et al., 2003). Following existing literature in cognitive science and psychology, we categorize the cognitive data as follows:

1. Gaze data

   - *Gaze Points:* Position of the eye-gaze on the screen
   - *Fixation:* A long stay of the gaze on a particular object on the screen. Fixations have both spatial (co-ordinates) and temporal(duration) dimensions
   - *Saccade:* A very rapid movement of the eyes between positions of rest. Forward (progressive) and backward (regressive) saccades are called *Progressions* and *Regressions* respectively
   - *Scanpath:* A line graph that contains fixations as nodes and saccades as edges

2. Key-stroke data

   - *Insertion:* Insertion of a word or a character inside running text
   - *Deletion:* Deletion of a word or a character
   - *Selection:* Highlighting a piece of text
   - *Rearrangement:* Dragging/Copy-pasting a piece of text from one position to another

3. Neuro-electromagnetic signals obtained through Electroencephalography (EEG), Magnetoencephalography (MEG) or brain-imaging obtained through functional Magnetic Resonance Imaging (fMRI).

For our research, we have relied on eye-tracking technology to capture and analyze cognitive information. While we do not undermine the power of other technologies (like EEG/MEG) in making rich cognitive information accessible, we believe eye-tracking can be useful in providing the first level insight into Cognitive NLP systems due to the following reasons:

- Technologies like brain imaging/EEG/MEG require a fairly complex and prohibitively expensive setup and hence, may not be used outside a laboratory. Multi-modal Cognitive NLP systems may consistently require cognitive data as input.

- Eye-tracking provides access to shallow cognitive information which is easily accessible, as inexpensive mobile eye-tracker are a reality now (Wood and Bulling, 2014; Yamamoto et al., 2013). This opens up avenues to get eye-tracking data from a large user-base non-intrusively.

- Key-logging can be done only when experiments involve typing. Most of the annotation tasks in NLP have reading as a major activity, limiting the possibilities of capturing cognitive data through key-stroke logging.

Human Eye-movement and Eye-tracking Technology

Human eye movement is poised between perception and cognition. The landing of eye-gaze on objects may seem to follow a random process, but it relates to the memory, expectations, and goals of the viewer. In general, the eye movement pattern during goal oriented reading is driven by- (a) perceptual properties of the text and (b) cognitive processes underlying language processing. Though the process seems biological, the role of eye-movement in the perception-action interplay can also be considered as a computational process from the perspective of *Computational Lens*[3], which enables us to see the computational side of various biological and natural processes like *learning in neural networks*,*response of immune system to an invading microbe*, *evolution of species*, etc. This has influenced psycholinguists and computational linguists to consider eye-tracking technology for cognitive studies related to natural language processing and understanding.

Eye-movement and Cognition in Reading

The relation between eye-movement attributes (such as fixations and saccades) and the corresponding cognitive process happening in the brain, especially during reading natural language text, is best explained through the *eye-mind hypothesis* (Just and Carpenter, 1980). The hypothesis is that when a subject views a word/object, he or she also processes it cognitively, for approximately the same amount of time he or she fixates on it. Though debatable (Anderson et al., 2004), the hypothesis has been considered useful in explaining theories associated with reading (Rayner and Duffy, 1986; Irwin, 2004; von der Malsburg and Vasishth, 2011). Gaze patterns are believed to indicate the conceptual difficulty the reader experiences (which, in turn, is linked with the cognitive effort (Sweller, 1988)). Linear and uniform-speed gaze movement is

[3] http://theory.cs.berkeley.edu/computational-lens.html

observed over texts having simple concepts, and often non-linear movement with non-uniform speed over more complex concepts (Rayner, 1998). This forms the basis of our research.

Eye-tracking Technology

Eye tracking refers to capturing the focus of a viewers gaze on a stimulus at a given time. This is typically done by tracking a viewers eye movements. Most commercially available eye-tracking systems today can capture eye movements using an unobtrusive method known as video-based corneal reflection (a detailed discussion is given in Chapter 2 of the thesis).

Eye-tracking is a century old technology and has seen tremendous advancements in the last few years, especially with the advent of high-quality eye-tracking devices. The technology has proven to be effective in solving important problems in the fields of neuroscience, industrial engineering, marketing, user experience design (refer to Duchowski (2002) for an overview).

Analyzing gaze data to gain insights into reading processes is a mature area of research (refer to Rayner (1998) for an overview). A number of successful models of eye-movement control for reading include the ones from Reichle and Laurent (2006), the E-Z Reader (Reichle et al., 2003, 2006), SWIFT (Engbert et al., 2005) and Bayesian inference based models (Bicknell and Levy, 2010; Engbert and Krügel, 2010). Eye-movement in reading has also been analyzed to study the correlation of eye-movement parameters derived from fixations and saccades with the lexical and syntactic complexities of text. Rayner and Duffy (1986) show how fixation time is associated with different forms of lexical complexity in the form of word frequency, verb complexity, and lexical ambiguity. Demberg and Keller (2008) relate complex eye-movement patterns to the syntactic complexity present in the text. von der Malsburg and Vasishth (2011) show that complex saccadic patterns (with higher degree of regression) are related to syntactic re-analysis arising from various forms of syntactically complex structures (*e.g.,* garden-path sentence).

Eye-tracking technology is a relatively new to NLP, with very few systems directly making use of gaze data in prediction frameworks. Doherty et al. (2010) show how eye tracking information can be used to evaluate output from Machine Translation systems. Martınez-Gómez and Aizawa (2013) use Bayesian Networks to model text-readability using eye-gaze information and linguistic properties of text Klerke et al. (2016) present a novel multi-task learning approach for sentence compression using labeled data, while, Barrett and Søgaard (2015) discriminate between grammatical functions using gaze features. These studies indicate the promise

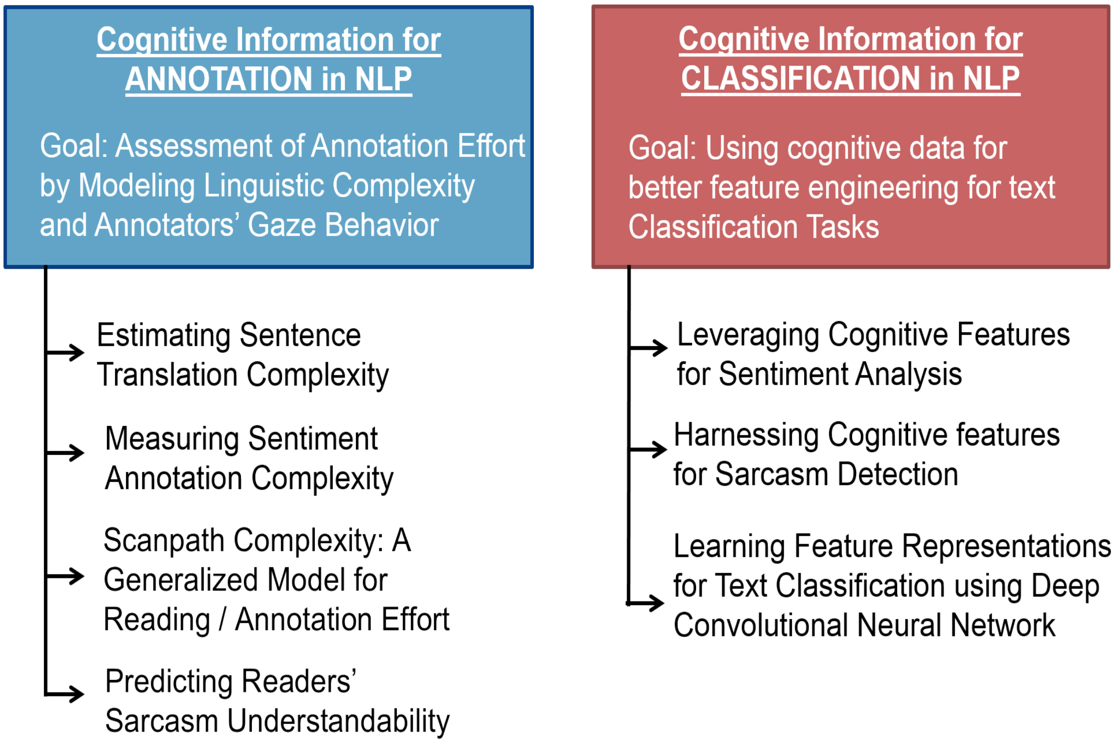

Figure 2: Research objectives and problems addressed

eye-tracking technology has in the field of linguistics and one can expect many large-scale experiments to be carried out in the near future.

Overview of Thesis Contribution

The thesis presents our research on incorporating cognitive modalities into NLP systems by harnessing cognitive information from human readers and annotators performing language processing tasks. We use *eye-tracking* technology to record and analyze *shallow cognitive information* during text *reading and annotation*, to gain insights into the cognitive underpinnings of difficult language processing tasks *viz.* Translation, Sentiment Analysis, and Sarcasm Detection. The insights gained through our analyses are translated into systems that help us to achieve the following objectives:

- Better assessment of **cognitive effort** in text-annotation, with a view to increasing human text-annotation efficiency.
- Extraction of **cognitive features**, with a view to improving the accuracy of text classifiers.

Our research objectives are summarized in Figure2 and are explained below.

Research Objective 1: Assessing Cognitive Effort in Text-annotation

As discussed earlier, annotation, being an integral part of data-driven NLP, can provide access to cognitive data. We record human reading activities during annotation by logging annotators' eye-movement patterns while they perform tasks like Translation and Sentiment Analysis. Gaze-data is modeled along with linguistic information to give rise to models of cognitive-effort for such annotation tasks. In this regard, we propose two cognitive models to estimate the complexity of text translation and sentiment analysis.

Estimating Text Translation Complexity

We introduce *Translation Complexity Index* (TCI), a measure of complexity in text translation. Realizing that any measure of TDI based on *direct* input by translators is fraught with subjectivity and ad-hocism, we rely on *cognitive evidence* from eye tracking. TCI is measured as the sum of *fixation (eye gaze)* and *saccade (eye movement)* times of the eye. We establish that TCI is correlated with various properties of the input sentence, like *length (L), the degree of polysemy (DP)* and *structural complexity (SC)*. We train a *Support Vector Regression* (SVR) system to predict TCIs for new sentences using these features as input. The primary purpose of this work is a way of ”binning” sentences (to be translated) in “easy,” “medium” and “hard” categories as per their predicted TCI. Our TCI metric is useful for proposing better cost-models for translation task, especially in a crowdsourcing scenario. This can also provide a way of monitoring progress of second language learners.

Measuring Sentiment Annotation Complexity of Text

Motivated by TCI, we follow a similar direction for sentiment annotation and propose a metric called Sentiment Annotation Complexity (SAC), a metric in sentiment analysis research that has been unexplored until now. First, the process of data preparation through eye tracking, labeled with the SAC score is formalized. Using our collected data set and a set of linguistic features, we train train a *Support Vector Regression* (SVR) system to predict SAC. Our predictive framework for SAC results in a mean percentage error of **22.02%**, and a moderate correlation of **0.57** between the predicted and observed SAC values. The merit of our work lies in (a) deciding the sentiment annotation cost in, for example, a crowdsourcing setting, (b) choosing the right classifier for sentiment prediction

Scanpath Complexity: A Generalized Model for Reading/Annotation Effort

For the previously discussed two problems, we rely on simplistic gaze-based measures like total fixation duration to label our data and later, successfully predict the labels using derivable textual properties. While measuring cognitive effort in annotation through total fixation/saccadde duration may seem robust under the assumption that "complex tasks require more time", it seems more intuitive to consider the complexity of eye-movement patterns in their entirety to while proposing such measures. This is under the assumption that complexity of gaze patterns are related to the cognitive effort required from the reader/annotator experiences (Sweller, 1988; Parasuraman and Rizzo, 2006; Just and Carpenter, 1980), which is linked to the underlying linguistic complexity.

We propose a cognitive model of measuring the complexity of eye-movement patterns (presented as scanpaths) of readers for text reading task; we call the measure *Scanpath Complexity*. Scanpath complexity is modeled as a function of various properties of gaze fixations and saccades- the basic parameters of eye movement behavior. We demonstrate the effectiveness of our scanpath complexity measure by showing that its correlation with different measures of lexical and syntactic complexity as well as standard readability metrics is better than the baseline measure of total duration of fixations. After grounding our scanpath complexity measure through reading experiments, we use similar approach to derive labels for Sentiment Annotation Complexity (SAC) prediction, engendering alternative systems for SAC.

Predicting Readers' Sarcasm Understandability

While performing the eye-tracking experiment for sentiment analysis, we observed that cognitive information could also be useful to model the ability of a reader to understand/comprehend certain aspects of a given reading material. This observation was quite clear in our sentiment annotation experiment (discussed in Chapter 3 of the thesis), where the eye-movement patterns of some of our annotators appeared to be subtle when the text had linguistic nuances like sarcasm, which the annotators could not recognize. This motivated us to work on a highly specific yet important problem of sarcasm understandability prediction - a potential starting point towards a more significant problem of modeling text comprehensibility.

Sarcasm understandability or the ability to understand textual sarcasm depends upon readers' language proficiency, social knowledge, mental state, and attentiveness. We introduce a novel method to predict the sarcasm understandability of a reader. The presence of *incongruity* in textual sarcasm often elicits distinctive eye-movement behavior by human readers. By recording and analyzing the eye-gaze data, we show that eye-movement patterns vary when sarcasm is understood *vis-à-vis* when it is not. Motivated by our observations, we propose a system for sarcasm understandability prediction using supervised machine learning. Our system relies on readers' eye-movement parameters and a few textual features, thence, is able to predict sarcasm understandability with an F-score of 93%. The availability of inexpensive embedded-eye-trackers on mobile devices creates avenues for applying such research which benefits web-content creators, review writers and social media analysts alike.

Research Objective 2: Extracting Cognitive Features for Text Classification

We consider eye movement data to extract *Cognitive Features*, to be used for text classification tasks *viz.*, Sentiment and Sarcasm detection along with traditional textual features. We propose two feature engineering schemes here - (1) Manually computing features from the text and gaze data by exploiting their underlying structure and (2) Automatically learning feature representations from the text and gaze inputs with the help of *Deep Neural Networks*. Our approaches that use these augmented text and gaze based feature-sets for classification consistently outperform state of the art sentiment and sarcasm classifiers that are based on text input alone. We observe that incorporation of cognitive features into traditional NLP systems can be useful in handling linguistic subtleties better, which accounts for the performance improvement.

Handcrafting Cognitive Features for Sentiment and Sarcasm Detection

Sentiment and sarcasm detection systems are often challenged by text with semantic and pragmatic subtleties. For example, the sentence *"I really love my job. I work 40 hours a week to be this poor."* requires an NLP system to be able to understand that the opinion holder has not expressed a positive sentiment towards her/his job. In the absence of explicit clues in the text, it is difficult for automatic systems to arrive at a correct classification decision, as they often lack external knowledge about various aspects of the text being analyzed. We try to tackle these subtleties by injecting cognitive information obtained through the eye-movement data of human annotators, in the form of features into traditional text-based classifiers. This is motivated by our observations from the eye-movement data which points to the fact that there exists a difference in eye-movement behavior between reading normal opinionated text and text with more linguistic subtleties. We augment traditional linguistic and stylistic features for sarcasm detection with the features obtained from readers eye movement data, extracted from simple eye-movement attributes and a graph structure underlying the gaze sequence. We perform statistical classification using the enhanced feature set so obtained. The augmented cognitive features improve the F-score of sentiment analysis by a maximum of 3.7% and 9.3% on two datasets and that of sarcasm detection by 3.7% , over the performance of the best-reported systems for the respective tasks.

Learning Features from Text and Gaze Inputs with Deep Neural Network

In this work, we contend that manual extraction of features is not good enough to tackle text subtleties that characteristically prevail in complex classification tasks like *sentiment analysis* and *sarcasm detection*, and that even the extraction and choice of features should be delegated to the learning system. We introduce a framework to automatically extract cognitive features from the *eye-movement/gaze* data of human readers reading the text and use them as features along with textual features for the tasks of sentiment polarity and sarcasm detection. Our proposed framework is based on deep Convolutional Neural Networks (CNN). The CNN *learns features* from both gaze and text and uses them to classify the input text. We test our technique on published sentiment and sarcasm labeled datasets, enriched with gaze information, to show that using a combination of automatically learned text and gaze features yields better classification performance over (i) CNN based systems that rely on text input alone and (ii) existing systems that rely on handcrafted gaze and textual features.

Dataset and Resources

All the datasets, resources for replicating the experiments and online tools are available at the Center for Indian Language Technology (CFILT)'s Cognitive NLP website (`http://www.cfilt.iitb.ac.in/cognitive-nlp`) and can be downloaded freely for academic use under Creative Common License.

Thesis Organization

We now provide a road map of the thesis which comprises 8 chapters. The broad theme of each work and the corresponding chapter/section in the thesis in which it is described is indicated.

Figure 3: Screenshot of the CFILT's Cognitive NLP website for downloading data

Chapter 1 is the introductory chapter of the thesis which gives an overview of the thesis. After motivating why cognitive information obtained through eye-tracking can be useful for Natural Language Processing, we give a brief background of eye-tracking technology and its applications. The objectives of our research are then briefly discussed.

Chapter 2 discusses various aspects of eye-tracking technology, summarizing the methods, theoretical aspects and applications of eye-tracking in NLP and other Areas. Section 2.1 is devoted to available eye-tracking systems and software and also provides an overview of the evolution of eye-tracking technology. Section 2.2 mentions domains other than NLP and psycholinguistics where eye-tracking has proven to be useful. Section 2.3 summarizes eye-movement research from the perspective of psycholinguistics and *reading* research. Sections 2.4 and 2.5 are the most relevant sections from the thesis's point of view as they discuss on how eye-tracking technology has helped to understand text-annotation process and improve NLP systems such as Part-of-speech Tagging, Machine Translation, and Sentence Compression.

The first part of our core work on *assessing cognitive effort in text-annotation by modeling linguistic complexity and annotators gaze behavior* is covered in Chapters 3, 4 and 5.

Chapter 3 describes how complexity of text-annotation can be modeled using gaze and text information for the tasks of translation and sentiment annotation. The first part of the chapter, *i.e.*, Section 3.1 proposes a method for estimating text translation complexity (TCI); the predictive framework for TCI is discussed in Section 3.1.3. Section 3.1.4 describes TCI as a function of translation processing time. Several linguistic and translation features used for TCI prediction are described in Sections 3.1.6-3.1.10. Section 3.1.11 discusses experiments and results for TCI prediction. The second part of the chapter, which is on measuring sentiment annotation complexity (SAC) of text begins in Section 3.2. Sections 3.2.4, 3.2.5 and 3.2.6 discuss the experimental set-up for the collection of eye-movement data, mathematical formulation of SAC, and linguistic and sentiment oriented features used for SAC prediction respectively. Sections 3.2.8 and 3.2.9 cover results and error-analysis.

Chapter 4 deals with *Scanpath Complexity*, introduced by us as a measure of cognitive effort for Text Reading/Annotation. Section 4.2 of the chapter discusses on how the complexity of eye-movement behavior (given in the form of scanpaths) can be mathematically modeled using several gaze attribute, which is discussed in section 4.3. The experiment setup is detailed in 4.4. Section 4.5 is devoted to detailed evaluation of scanpath complexity. Finally, how a measure like scanpath complexity can be useful in annotation settings is discussed in Section 4.7.

Chapter 5 is on predicting readers' sarcasm understandability by modeling gaze behavior. In this regard, our hypothesis on how cognitive processes behind sarcasm comprehension are related to eye-movement behavior is explained in Section 5.2. Sections 5.3 and 5.4 are devoted to creation and analysis of eye-tracking data for sarcasm understandability respectively. Section 5.5 explains the predictive framework followed by a detailed discussion of results

The second part of our core work on *leveraging cognitive data for improving NLP systems* is covered in Chapters 6 and 7.

Chapter 6 introduces a novel feature engineering scheme for short-text classification. The first part pf the chapter, *i.e.*, Section 6.1 explains how cognitive data can be leveraged in the form of features for the task of sentiment analysis. After setting the motivation, we describe the dataset used for experimentation and performance of existing text-based classifiers in Sections 6.1.3 and 6.1.4. Section 6.1.5 introduces our novel gaze-based feature design. Experiments and results are covered in Section 6.1.6 and the importance of the cognitive features is examined in Section 6.1.7. The second part of the chapter is on sarcasm classification using cognitive features derived from eye-movement data, which follows a very similar approach as the first part and is detailed in Section 6.2. The final section of this chapter, *i.e.*, Section 6.3 discusses how feasible it is to collect and use eye-movement information for NLP systems.

Chapter 7 introduces a novel deep Convolutional Neural Network (CNN) based framework for automatically learning cognitive features from the gaze and text data for sentiment and sarcasm classification. Section 7.3 discusses the motivation behind preferring deep learning based frameworks for feature extraction over manual feature engineering. Section 7.4 further motivates the idea behind choosing CNNs over other available alternatives for feature extraction and classification. The CNN architecture is proposed and discussed in section 7.5. Section 7.6.1 refers to the publicly available dataset used for our experimentation whose set-up details are given in section 7.6. We discuss our results in section 7.7 and provide a detailed analysis of the results along with some more insightful observations in section 7.8.

The thesis is concluded in Chapter 8 with relevant pointers to future directions and possible implications.

Bibliography

Agirre, E. and Rigau, G. (1996). Word sense disambiguation using conceptual density. In *Proceedings of the 16th conference on Computational linguistics-Volume 1*, pages 16–22. Association for Computational Linguistics.

Anderson, J. R., Bothell, D., and Douglass, S. (2004). Eye movements do not reflect retrieval processes limits of the eye-mind hypothesis. *Psychological Science*, 15(4):225–231.

Antonenko, P., Paas, F., Grabner, R., and van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22(4):425–438.

Barrett, M. and Søgaard, A. (2015). Using reading behavior to predict grammatical functions. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pages 1–5, Lisbon, Portugal. Association for Computational Linguistics.

Bicknell, K. and Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the 48th annual meeting of the ACL*, pages 1168–1178. ACL.

Demberg, V. and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Doherty, S., OBrien, S., and Carl, M. (2010). Eye tracking as an mt evaluation technique. *Machine translation*, 24(1):1–13.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470.

Engbert, R. and Krügel, A. (2010). Readers use bayesian estimation for eye movement control. *Psychological Science*, 21(3):366–371.

Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). Swift: a dynamical model of saccade generation during reading. *Psychological review*, 112(4):777.

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. *The interface of language, vision, and action: Eye movements and the visual world*, pages 105–134.

Joshi, S., Kanojia, D., and Bhattacharyya, P. (2013). More than meets the eye: Study of human cognition in sense annotation. naacl hlt 2013. *Atlanta, USA*.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329.

Klerke, S., Goldberg, Y., and Søgaard, A. (2016). Improving sentence compression by learning to predict gaze. *arXiv preprint arXiv:1604.03357*.

Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.

Martınez-Gómez, P. and Aizawa, A. (2013). Diagnosing causes of reading difficulty using bayesian networks. *IJCNLP 2013*.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, 41(2):10.

Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1):63–71.

Parasuraman, R. and Rizzo, M. (2006). *Neuroergonomics: The brain at work*. Oxford University Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Rayner, K. and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201.

Reichle, E. D. and Laurent, P. A. (2006). Using reinforcement learning to understand the emergence of" intelligent" eye-movement behavior during reading. *Psychological review*, 113(2):390.

Reichle, E. D., Pollatsek, A., and Rayner, K. (2006). E–z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7(1):4–22.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The ez reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, 26(04):445–476.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285.

von der Malsburg, T. and Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65(2):109–127.

Wood, E. and Bulling, A. (2014). Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 207–210. ACM.

Yamamoto, M., Nakagawa, H., Egawa, K., and Nagamatsu, T. (2013). Development of a mobile tablet pc with gaze-tracking function. In *Human Interface and the Management of Information. Information and Interaction for Health, Safety, Mobility and Complex Environments*, pages 421–429. Springer.