Towards Predicting Age of Acquisition of Words Using a Dictionary Network

Ditty Mathew*, **Girish Raguvir Jeyakumar***, **Rahul Kejriwal***, **Sutanu Chakraborti**

Department of Computer Science and Engineering
Indian Institute of Technology Madras
Chennai 600 036, India
ditty@cse.iitm.ac.in
{girishraguvir, kejriwalrahul.1996}@gmail.com
sutanuc@cse.iitm.ac.in

Abstract

Age of Acquisition (AoA) of words is an important psycho-linguistic property that influences various lexical tasks such as speed of reading words, naming pictures etc. In this paper, we study the effectiveness of graph theoretic and lexical features derived from dictionary networks in predicting the Age of Acquisition of English words. We show that dictionary networks contain a lot of information that can hint at the AoA of words, and the result promises that there are significant improvements over earlier approaches. We also extended the experiment to predict the Age of Acquisition of Hindi words and evaluated using words in Hindi textbooks.

1 Introduction

Age of Acquisition (AoA) (Kuperman et al., 2012) refers to the age at which a word is typically learned. For example, 'penguin' is generally learned earlier than 'albatross' and hence, 'penguin' has a lower age of acquisition value. Age of acquisition is an important feature for lexical decision tasks (Gilhooly and Logie, 1982) like speed of word reading, picture naming, word retrieval from lexical memory etc. AoA is important for two reasons. Firstly, word frequency is used as a feature alongside AoA in psycho-linguistic studies and is estimated from corpora consisting of materials meant for adult readers (Gerhand and Barry, 1999). Thus, word frequency typically underestimates the importance of words acquired earlier and thus, AoA plays an important role in complementing the information provided by word frequency. Secondly, it is believed that words acquired earlier have internal representations that can be activated faster independent of the number of times the word has been encountered. Despite the importance of the AoA parameter, the only way to estimate its value for any word is by surveying human participants. However, this process is in general tedious, requires active human participation and is thus limiting. We explore the possibility of an easier alternative. For this, we utilize the structure of word dictionary (Picard et al., 2013) where words are constructively defined.

Dictionaries provide recursive definitions and establish a dependency relation between words. To observe the inherent notion of order in which words are acquired, we consider a network constructed from a dictionary, which is a directed graph with words as vertices and edges denoting definitional participation. An edge from vertex A to vertex B implies that A is used in the definition of B in the dictionary. Thus, all predecessors of a vertex must be known for the vertex to be understood by a reader. The inherent idea is that some words (like "green") have to be learned via sensorimotor experience (Harnad, 1990), hence they are expected to be involved in loops/cycles in the network, while others can possibly be learned from constructive definitions composed of other known words which are probably simpler (Miller et al., 1990). Picard et al. (2013) extensively studied the hidden structure of the dictionary and used AoA to discriminate between different dictionary network regions successfully. This provides a strong motivation to explore the dictionary network structure for our problem - to estimate the AoA of words. In this paper, we build on this to take a step towards a faster alternative that can approximate AoA value for any word based on the dictionary network, and thus overcome the limited availability of data in addition to alleviating the need for tedious surveys.

* denotes equal contribution

Proc. of ICON-2018, Patiala, India. December 2018 ©*2018 NLPAI*, pages 193–201

It is important to keep in mind that the AoA is often highly subjective and can widely vary among subjects depending on their educational background, mental capabilities etc. So predicting AoA is in some way an ill-defined problem with an inherent noise in the data. Thus the primary motivation of the paper is to reveal the extent to which we can successfully predict AoA while restricting our scope to features derived from dictionary networks. It should be noted that aligning AoA may need richer cognitive and psycholinguistic features that are not contained in the dictionary network.

2 Background

2.1 Structure of Dictionary Network

Massé et al. (2008) developed a formal groundwork to determine "how word meaning is explicitly grounded in real dictionaries" and observed that meaning cannot be fetched based on dictionary definition recursively; at some point circularity of definitions must be broken by grounding the meaning of certain words. Massé et al. (2008) proposed a greedy algorithm to obtain a set of words in the dictionary network from which the rest of the words in the network can be defined without any circular dependencies. This set of words is called *grounding kernel*. The grounding kernel is estimated by repeatedly removing the nodes that do not have out-neighbors until there are no such nodes in the network. The intuition is that the nodes which do not have out-neighbors are not used for defining other words, so these words do not lead to circularity problem.

Picard et al. (2013) analyzed the full structure of the dictionary network, and they have found that the grounding kernel of dictionaries consist of a set of words, approximately 10% of the size of the entire dictionary, from which all other words can be defined. Inside the kernel, two sets of words are identified. One set is called *core* which includes the words in the strongly connected components (SCCs) that act as sources, i.e., all the words in these SCCs are defined by words in that SCC itself (in graph theoretical terms the nodes corresponding to such SCCs do not have incoming edges in its SCC-condensed graph[1]). The second set is called *satellites* which include rest of the SCCs in the grounding kernel.

[1]Each strongly connected component of a graph will be condensed to a single node in its SCC-condensed graph.

Vincent-Lamarre et al. (2016) studied the hierarchies of concepts in the concept network and proposed two types of hierarchies - kernel hierarchy and core hierarchy. These hierarchies are based on a graph theoretic property called *definitional distance*. In kernel hierarchy, the definitional distance of a word u is defined as,

1. $dist(u) = 0$, if $u \in kernel$
2. $dist(u) = 1 + \max\{dist(v) : v \in predecessors(u)\}$, otherwise

The words with definitional distance, $dist = 0$ are the words in the zeroth level of the kernel hierarchy. All words in the kernel of a dictionary will be at zeroth level. The words, which can be defined using the words in the zeroth level, are in the first level of the hierarchy and so on. For the words in the i^{th} level of the hierarchy, all the words which define these words should be at any level between 0 and $i-1$.

In core hierarchy, the definitional distance of words are computed with respect to core where core is a set of strongly connected components. The core hierarchy is a hierarchy of strongly connected components. The definitional distance of a word u from core is defined as,

1. $dist(u) = 0$, if $u \in core$
2. $dist(u) = 1 + \max\{dist(v) : \exists w \in SCC(u)$ such that $v \in predecessors(w)\}$, otherwise

where $SCC(u)$ contains all nodes in the strongly connected component that contains u.

2.2 Study of Psycho-linguistic Properties based on Dictionary Network

Vincent-Lamarre et al. (2016) studied the psycholinguistic properties such as frequency, concreteness, and age of acquisition of words with respect to the dictionary structure. This study is performed in dictionaries such as Cambridge dictionary, Longman dictionary, Merriam Webster dictionary and WordNet. Frequencies of words are computed using SUBTLEX (US) corpus; the concreteness ratings for 40,000 English word lemmas given in (Brysbaert et al., 2014) are used for concreteness; the age of acquisition ratings for 30,000 English words (Kuperman et al., 2012) are used for the age of acquisition property. They observed that the average age of acquisition of words that are part

of the core is smaller than the average age of acquisition of words that are part of satellites which is in turn smaller than the average age of acquisition of words that are not part of core and satellites. A similar trend is observed for frequency of words. For concreteness, the average concreteness value of words in satellites is observed to be smaller than the average concreteness value of words that are part of the core, which is in turn smaller than that of rest of the words. The properties such as frequency and age of acquisition follow the similar observation and it is hypothesized that words in the core are more frequent and acquired earlier compared to satellites, and words in the rest of the dictionary are less frequent and acquired later compared to satellites. This motivates us to test whether the dictionary structure will help in predicting the age of acquisition of words.

3 Dictionary Network Based Features

Let $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ be the directed graph constructed using words in the dictionary D. Each vertex $v \in \mathbb{V}$ indicates a word and each directed edge $(u, v) \in \mathbb{E}$ represents that the word u is used to define the word v in dictionary D. We propose graph-based features based on dictionary network $\mathbb{G}$ to predict the age of acquisition of each word defined in D.

We use the following basic dictionary structure-based features which were analyzed by Vincent-Lamarre et al. (2016) as having patterns related to AoA.

1. **Is core**: This is a binary value, which indicates whether the word is part of core or not. This feature is based on the work by Vincent-Lamarre et al. (2016) where they show that the words in the core are learned earlier than the satellite words, which are in turn learned earlier than the rest-of-dictionary.

2. **Is kernel**: This feature indicates whether the word is part of the kernel or not. This feature is motivated by the observation that the words in the kernel are acquired earlier than the words in the rest-of-dictionary (Vincent-Lamarre et al., 2016).

3. **Definitional distance from kernel and core**: These two features are defined by Vincent-Lamarre et al. (2016). For both definitional distances from core and kernel, Vincent-Lamarre et al. (2016) observed a linear trend between the definitional distance and the average age of acquisition of words at each definitional distance.

We propose the following features which are derived from the basic dictionary features studied by Vincent-Lamarre et al. (2016).

1. **PageRank**: The out-neighbors of a word w in $\mathbb{G}$ are the words which are defined using w. One would expect w to be acquired before words that are defined using w. Thus, *a word is acquired early if it is used to define several words that are acquired early; this leads to a circularity*. In order to resolve this circularity, we use the PageRank (Page et al., 1999) of words in the transpose graph $\mathbb{G}'$ whose vertices are same as $\mathbb{G}$, but edges are reversed to capture the importance from out-neighbors in $\mathbb{G}$.

2. **SCC PageRank**: The words within the strongly connected components (SCC) are closely associated. Because of circular dependencies between words in SCCs, it is hard to find which words are defined first. The correlation between definitional distance from core and AoA showed by (Vincent-Lamarre et al., 2016) claims that if there is an edge from the condensed node of SCC S_j to the condensed node of SCC S_i in the condensed graph, the words in S_i would be acquired earlier than words in S_j. In order to analyze the random walk interpretation of this property, we consider the importance of the SCC to which each word is associated as a feature and it is estimated as the PageRank of the corresponding condensed node in the condensed graph.

3. **Within SCC PageRank**: This feature computes the importance of a word within the SCC it belongs to as given by PageRank. We consider this feature to complement the SCC PageRank feature.

We propose the following local features to understand their trends in dictionary network for AoA prediction task.

1. **Word length**: The number of characters in the word is used as a feature. Generally, at lower ages, shorter words are learned and longer words tend to be acquired at later ages.

2. **In-degree centrality**: The in-degree centrality of a word is computed by dividing its in-degree by maximum possible in-degree. In-degree of a word w is a measure of the number of other words that are used in the definition of w. The intuition behind this is that a larger value suggests that this word may be acquired later.

3. **Out-degree centrality**: The out-degree centrality is the ratio of the out-degree of a word to the maximum possible out-degree. The out-degree of a word w is a measure of the number of words that are defined using w. The intuition is that the larger value should mean that this word may be acquired earlier.

4. **Local clustering coefficient**: The local clustering coefficient is the number of edges between immediate predecessors and successors divided by the maximum number of possible edges among them. This feature is used to study the effect of clustering tendencies on the AoA of a word.

5. **Second in-neighborhood and out-neighborhood**: The number of predecessors at distance 1 or 2 from a word in the graph is taken as the second in-neighborhood. Similarly, second out-neighborhood is the number of successors at distance 1 or 2 from a word in the graph. These features are used to study the higher level significance of in-degree centrality and out-degree centrality.

4 Experimental Evaluation

4.1 Datasets

We use the age of acquisition data from the Kuperman's (Kuperman et al., 2012) dataset and MRC psycho-linguistic dataset (Coltheart, 1981). The Kuperman dataset[2] contains AoA values for 31,124 words which are collected using Amazon Machine Turk. It contains data aggregated by asking participants to give one value in the range 1 to 25 for each word. The final AoA for a given word is then computed by taking the average of all the responses. The MRC psycho-linguistic dataset[3] contains lexical, morphological and psycho-linguistic properties of 1,50,837 words out of which the age of acquisition of 1,903 words are available. We use the age of acquisition ratings from both Kuperman and MRC datasets for evaluation.

We construct dictionary networks from Cambridge International Dictionary of English (CIDE), Longman Dictionary of Contemporary English (LDOCE), Merriam-Webster (MWC) dictionary and WordNet. We use the first definition of first sense in these dictionaries to build the corresponding dictionary graph. The stop words in the word definition are removed. The number of words in the network that is constructed from all four dictionaries are given in Table 1.

Dictionary	No of Words
CIDE	19,614
LDOCE	26,859
MWC	79,979
WordNet	76,792

Table 1: No of words in the network constructed from all four dictionaries

The graph-based features are extracted from these dictionary networks for the words that are common in all dictionaries and AoA dataset. There are 14,436 common words for Kuperman dataset, and we obtain 1,495 common words for MRC dataset.

4.2 Experiment Setup

We propose to use the dictionary network-based features along with richer cognitive and psycho-linguistic features for the prediction of age of acquisition of words. Hence, we use dictionary network-based features along with the lexical features and semantic features proposed in Paetzold and Specia (2016a) to predict the age of acquisition of words. The lexical features include

- Number of syllables
- Word's frequency in the Brown (Francis and Kucera, 1979), SUBTLEX (Brysbaert and New, 2009), SubIMDB (Paetzold and Specia, 2016b), Wikipedia, Simple Wikipedia (Kauchak, 2013) corpora
- Number of senses, synonyms, hypernyms, and hyponyms for word in WordNet

[2]Kuperman dataset is downloaded from http://crr.ugent.be/archives/806

[3]MRC psycho-linguistic dataset is downloaded from http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm

- Minimum, maximum and average distance between the word's senses in WordNet and the root sense
- Number of images found for word in the Getty Image database[4]

The semantic features are the word embedded vectors (Mikolov et al., 2013) of words. The word embedded vectors capture the semantic information of words.

We train a ridge regression model to predict the AoA of words and train the model using lexical, semantic and dictionary features. We train the model using different sets of features,

1. Lexical features (Paetzold and Specia, 2016a)
2. Lexical and Semantic features (Paetzold and Specia, 2016a)
3. Lexical, Semantic and Dictionary network features based on CIDE dictionary
4. Lexical, Semantic and Dictionary network features based on LDOCE dictionary
5. Lexical, Semantic and Dictionary network features based on MWC dictionary
6. Lexical, Semantic and Dictionary network features based on WordNet

We use the lexical features and semantic features proposed in Paetzold and Specia (2016a) as a baseline to predict the age of acquisition of words. We compare the baseline model with the model that uses lexical features, semantic features, and dictionary features. We used the word embedded vectors trained using Google news dataset[5].

The model is evaluated by analyzing the Spearman's (ρ) (Spearman, 1906), Pearson's (r) (Pearson, 1920) and Kendall's tau (Sen, 1968) correlation coefficients between the actual AoA ranking and the predicted AoA ranking of words in test data. The Spearman's correlation is a measure of rank correlation and it assesses the monotonic relationships between variables. The Pearson's correlation measures the linear correlation between two variables. The Kendall's tau coefficient measures the association of variables based on pairwise ordering.

[4] http://developers.gettyimages.com/

[5] https://code.google.com/archive/p/word2vec/

4.3 Evaluation

We analyze the correlation coefficients on a 10-fold train-test splits and the average Spearman's (ρ), Pearson's (r), Kendall's tau correlations of test data are obtained. We compare the models which use dictionary network-based features with the baselines which do not use dictionary network-based features. The correlation coefficients of the predicted value with respect to the Kuperman dataset for all feature sets are listed in Table 2. All correlation coefficients are statistically significant with $p < 0.05$.

Features	Avg Spearman	Avg Pearson	Avg Kendall's tau
Lexical Features (Paetzold and Specia, 2016a)	0.4837	0.5049	0.3353
Lexical + Semantic Features (Paetzold and Specia, 2016a)	0.7793	0.7835	0.5856
Lexical + Semantic + CIDE	**0.7926***	**0.8004***	**0.5986***
Lexical + Semantic + LDOCE	0.7910*	0.7990*	0.5970*
Lexical + Semantic + MWC	0.7837*	0.7887*	0.5899*
Lexical + Semantic + WordNet	0.7878*	0.7924*	0.5941*

Table 2: Correlation Coefficients between the predicted AoA of words and the AoA based on Kuperman dataset; * indicates improvements are statistically significant with $p < 0.05$

We can observe that the model which uses dictionary network features result in better correlations compared to both baselines. The improvements in correlations are statistically significant with $p < 0.05$.

The correlation coefficients between the predicted AoA of words and the age of acquisition values in MRC psycho-linguistic dataset using all sets of features are given in Table 3. The correlation coefficients obtained when trained with all features are consistently performing better in all four dictionaries compared to baseline models.

Some words which are predicted as acquired earlier compared to its actual AoA using lexical and semantic (Paetzold and Specia, 2016a) features have their AoA predicted better when dictionary features are used. These words are observed as non-kernel words in the dictionary network. We also observed that the AoA prediction error is very less for words with a lower age of acquisition such as 'give', 'work', 'show', 'day' etc. when dictionary features are used compared to the baseline features. The inclusion of lexical and semantic features improved the prediction value for

Features	Avg Spearman	Avg Pearson	Avg Kendall's tau
Lexical Features (Paetzold and Specia, 2016a)	0.5632	0.5514	0.3993
Lexical + Semantic Features (Paetzold and Specia, 2016a)	0.7576	0.7700	0.5718
Lexical + Semantic + CIDE	0.7867*	**0.7957***	**0.6008***
Lexical + Semantic + LDOCE	**0.7880***	0.7956*	0.6005*
Lexical + Semantic + MWC	0.7609	0.7713	0.5746
Lexical + Semantic + WordNet	0.7703*	0.7820*	0.5830*

Table 3: Correlation Coefficients between the predicted AoA of words and the AoA based on MRC psycho-linguistic dataset; * indicates improvements are statistically significant with $p < 0.05$

Features	Avg Spearman	Avg Pearson	Avg Kendall's tau
Lexical Features (Paetzold and Specia, 2016a)	0.5673	0.4956	0.4004
Lexical + Semantic Features (Paetzold and Specia, 2016a)	0.7670	0.7358	0.5707
Lexical + Semantic + CIDE	0.7987*	**0.7672***	0.6026*
Lexical + Semantic + LDOCE	**0.8042***	0.7654*	**0.6068***
Lexical + Semantic + MWC	0.7741*	0.7467*	0.5783*
Lexical + Semantic + WordNet	0.7785*	0.7572*	0.5836*

Table 4: Correlation Coefficients between the predicted familiarity of words and the familiarity rating based on MRC psycho-linguistic dataset; * indicates improvements are statistically significant with $p < 0.05$

words which are i) general words like 'something', 'everyday', ii) colloquial words like 'grandma', 'mama', 'papa'. This is because of the infrequent use of such words in dictionary definitions.

We also study the prediction of other psycho-linguistic properties using dictionary network-based features. We experimented with the psycho-linguistic properties such as familiarity and imagability (Gilhooly and Logie, 1980). The familiarity of a word is the frequency with which a word is seen, heard and used. The imagability of a word is the intensity with which a word arouses images. We use the familiarity and imagability ratings presented in MRC psycho-linguistic dataset (Coltheart, 1981). This dataset contains familiarity ratings for 3,814 words and imagability ratings for 3,733 words. We trained a ridge regression model with the target value as the familiarity rating for predicting the familiarity of a word and the imagability rating is used as the target value for predicting the imagability of a word.

The same set of features are used for comparison and observed that both familiarity and imagability ratings are correlating better when dictionary network based features are used. All correlation coefficients are statistically significant with $p < 0.05$. The improvements in correlations are also statistically significant with $p < 0.05$. The dictionary features based on CIDE and LDOCE dictionaries are performing better than other two dictionaries.

Features	Avg Spearman	Avg Pearson	Avg Kendall's tau
Lexical Features (Paetzold and Specia, 2016a)	0.4734	0.4681	0.3252
Lexical +Semantic Features (Paetzold and Specia, 2016a)	0.7829	0.7708	0.5857
Lexical +Semantic + CIDE	**0.7915***	**0.7807***	**0.5939***
Lexical +Semantic + LDOCE	0.7912*	0.7797*	0.5936*
Lexical +Semantic + MWC	0.7860*	0.7757*	0.5877*
Lexical +Semantic + WordNet	0.7890*	0.7785*	0.5914*

Table 5: Correlation Coefficients between the predicted imagability of words and the imagability rating based on MRC psycho-linguistic dataset; * indicates improvements are statistically significant with $p < 0.05$

4.4 Experiment using School Textbooks

In this experiment, our task is to predict the class or grade[6] at which a word can be introduced in the school curriculum. We used the words present in Indian English textbooks from standard 1 to standard 10 which are published by National Council of Educational Research and Training (NCERT)[7]. If a word is first mentioned in i^{th} standard where $1 \leq i \leq 10$, then the target value of that word is i. In this way, we labeled the target value of words that are used in standard 1 to 10 English textbooks. We extracted all words from these textbooks and out of which we used the 5,496 words that are common in all four dictionaries.

For this task, we prefer regression over classification as the target values are ordered. We trained

[6] we use the words 'class', 'grade', or 'standard' interchangeably.

[7] http://www.ncert.nic.in/ncerts/textbook/textbook.htm

a ridge regression model to predict the target value and evaluated the model using mean squared error over 10-fold train-test splits. The model is trained using all sets of features and the results are given in Table 6. The mean squared error is highest when trained using only lexical features and it is decreased when lexical and semantic features are used. The mean squared error is even reduced when dictionary network-based features are used along with lexical and semantic features. The error is minimum when dictionary network features from LDOCE dictionary are used. The reduction in mean squared error is statistically significant with $p < 0.05$ when dictionary network-based features from CIDE, LDOCE and WordNet dictionaries are used and the deduction is statistically significant with $p < 0.01$ when MWC dictionary based features are used.

Features	Mean Squared Error
Lexical Features (Paetzold and Specia, 2016a)	0.8039
Lexical + Semantic Features (Paetzold and Specia, 2016a)	0.6621
Lexical + Semantic + CIDE	0.6326*
Lexical + Semantic + LDOCE	**0.6289***
Lexical + Semantic + MWC	0.6595+
Lexical + Semantic + WordNet	0.6553*

Table 6: Average Mean squared error when predicted the standard at which a word is first introduced to school students in NCERT English textbooks; * indicates deduction in mean squared error is statistically significant with $p < 0.05$ and + indicates deduction is statistically significant with $p < 0.01$

We also extended our experiment for Hindi words by using Hindi textbooks published by NCERT. Similar to the experiment using English Words, here the task is to predict the class or standard at which a Hindi word can be introduced in the school curriculum. We used the gloss of Hindi words in Hindi WordNet (Bhattacharyya, 2017) to construct the dictionary network using which the network based features are extracted. For lexical features, we used the frequency of words based on Hindi Corpus[8] published by Center for Indian Language Technology, IIT Bombay. The Hindi WordNet is used for other lexical features based on WordNet. We used Hindi Wikipedia[9] dump to train the embedded vectors of Hindi words. We used these features to train a ridge regression model to predict the standard at which a Hindi word can be introduced. We extracted words from NCERT Hindi textbooks from standard 1 to 10 and all features are obtained for 3,860 words. The model is evaluated using mean squared error over 10-fold train-test splits. The average mean squared error obtained when trained using i) only lexical features, ii) lexical and semantic features, iii) lexical, semantic and dictionary network features are given in Table 7. We can observe that the addition of dictionary network features improves the prediction compared to lexical and semantic features. This experiment also signifies the impact of the proposed model in predicting the age of acquisition of Hindi words for which an AoA dataset is not available.

Features	Mean Squared Error
Lexical Features (Paetzold and Specia, 2016a)	0.8708
Lexical + Semantic Features (Paetzold and Specia, 2016a)	0.8056
Lexical + Semantic + Dictionary	**0.7674***

Table 7: Average Mean squared error when predicted the standard at which a word is first introduced to school students in NCERT Hindi textbooks; * indicates deduction in mean squared error is statistically significant with $p < 0.05$

Class	Words
1	साथ (saath), काम (kaam), याद (yaad), फूल (phuul), जानवर (jaanvar)
2	आराम (aaram), दौरान (dauran), घास (ghas), असर (asar), पुलिस (pulice)
3	देखभाल (dekhbhaal), मुख (mukh), झलक (jhalak), वार (vaar)
4	दिमाग (dimaag), ट्रेन (train), जीव (jeev), उपहार (upahaar), खैर (khair)
5	पकवान (pakvaan), नौकर (naukar), बरतन (bartan), नमकीन (namkeen), मांग (maang)
6	जूट (juut), ढोलक (dholak), अधिगम (adhigam), आंगन (aangan), आश्वासन (aaswaasan)
7	गर्जन (garjan), खड्ग (khadag), जुल्म (julm), विकर्ण (vikarn), सूर्यदेव (sooryadev)
8	प्रेरक (prerak), आत्मसम्मान (aatmasamman), समाहित (samaahit), प्रतिमान (praratimaan)
9	अवमूल्यन (avamuulyan), वैष्णव (vaishnav), जागीरदार (jaageeradhaar), अध्यात्म (adhyaathm)
10	पाठ्यपुस्तक (paathyapustak), पूँजीवाद (punjeevaad), मूर्तिकार (muurtikaar), विद्याधर (vidhyaadhar)

Table 8: Examples of Hindi words predicted for Class/Standard 1 to 10

In order to qualitatively analyze the results, the predicted value of words is rounded when trained using all features. The class at which a word can be introduced is correctly predicted when the rounded predicted value is the same as the actual value assigned to the word. Some examples of Hindi words with the class/standard are correctly predicted are

[8] http://www.cfilt.iitb.ac.in/Downloads.html
[9] https://hi.wikipedia.org/

given in Table 8.

5 Discussion

Vajjala and Meurers (2014) proposed psycho-linguistic features such as the age of acquisition, familiarity, imagability and concreteness for predicting the reading level of a text and it is used for assessing the relative reading level of sentence pairs for text simplification. Our experiments suggest that the dictionary network-based features can be used as lexical features for predicting the reading level of a text.

In our experiments, the dictionary network-based features are performing better when extracted from Cambridge (CIDE) and Longman (LDOCE) dictionaries compared to Merriam Webster (MWC) and WordNet dictionaries. The motivation of proposing dictionary network-based features for predicting age of acquisition is the observation by Vincent-Lamarre et al. (2016) that the average age of acquisition of words in the core is smaller than the average age of acquisition of words in the satellites which is in turn smaller than the average age of acquisition of words in the rest of the dictionaries. In Figure 1, the average age

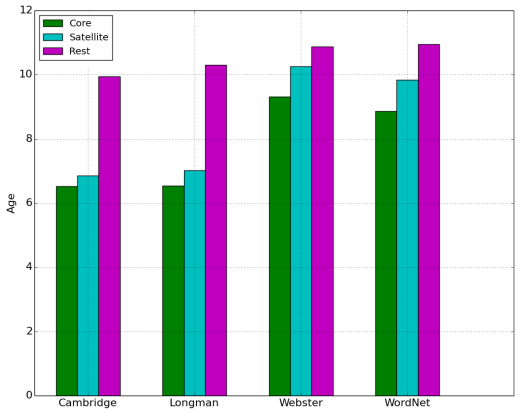

Figure 1: Average age of acquisition of words in Core, Satellites and Rest of the words in all four dictionaries

of acquisition values of words in core, satellites and the rest of the words are given for all four dictionaries. We can observe that the average AoA values are increasing from core to satellites to rest of the words in all four dictionaries. The average AoA of words in the kernel is the average of the AoA of words in core and satellites. From Figure 1, it is clear that the average AoA of kernel words will be smaller than the rest of the words. We can observe that the differences between the average AoA value of core words and the rest of the words are small in MWC and WordNet dictionaries, whereas the differences are large in CIDE and LDOCE dictionaries. This may be the reason that the CIDE and LDOCE dictionaries are performing better than MWC and WordNet dictionaries.

Concreteness (Brysbaert and New, 2009) is another psycho-linguistic property which is widely used along with age of acquisition for reading level prediction, text simplification etc. Concreteness is the extent to which the object that the word can be experienced by senses. The proposed dictionary network features are based on the dictionary structure where core words are defined before satellite words which are in turn defined before the rest of the words in the dictionary. Vincent-Lamarre et al. (2016) observed that the words present in the satellites are more concrete than the words present in the core. Since concreteness does not follow the progression from core to satellite and then to the rest of the words, which is central to our hypothesis, we did not use dictionary features to estimate concreteness.

6 Conclusion

We study the effectiveness of graph-theoretic features derived from dictionary networks in predicting the age of acquisition of words and the result shows significant improvements over earlier approaches that relate dictionary features to AoA. This work is a step towards understanding the difficult cognitive task of understanding how we acquire words. We also study the usefulness of dictionary network-based features for predicting other psycho-linguistic properties such as familiarity and imagability and the results are promising.

To the best of our knowledge, the psycho-linguistic study on Hindi has not been done before. Our experiment on Hindi words signifies the impact of the proposed model in predicting the age of acquisition of Hindi words.

References

Pushpak Bhattacharyya. 2017. IndoWordNet. In *The WordNet in Indian Languages*, pages 1–18. Springer.

Marc Brysbaert and Boris New. 2009. Moving beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*, 41(4):977–990.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas. *Behavior Research Methods*, 46(3):904–911.

Max Coltheart. 1981. The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology*, 33(4):497–505.

WN Francis and H Kucera. 1979. Brown Corpus Manual. *Brown University*.

Simon Gerhand and Christopher Barry. 1999. Age of Acquisition, Word Frequency, and the Role of Phonology in the Lexical Decision Task. *Memory & Cognition*, 27(4):592–602.

Ken J Gilhooly and Robert H Logie. 1980. Age-of-acquisition, Imagery, Concreteness, Familiarity, and Ambiguity Measures for 1,944 Words. *Behavior Research Methods & Instrumentation*, 12(4):395–427.

KJ Gilhooly and RH Logie. 1982. Word Age-of-Acquisition and Lexical Decision Making. *Acta Psychologica*, 50(1):21–34.

Stevan Harnad. 1990. The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

David Kauchak. 2013. Improving Text Simplification Language Modeling using Unsimplified Text Data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1537–1546.

Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. Age-of-acquisition Ratings for 30,000 English Words. *Behavior Research Methods*, 44(4):978–990.

A Blondin Massé, Guillaume Chicoisne, Yassine Gargouri, Stevan Harnad, Olivier Picard, and Odile Marcotte. 2008. How is Meaning Grounded in Dictionary Definitions? In *Proceedings of the 3rd Textgraphs Workshop on Graph-Based Algorithms for Natural Language Processing*, pages 17–24. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–244.

Gustavo Paetzold and Lucia Specia. 2016a. Inferring Psycholinguistic Properties of Words. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 435–440.

Gustavo H Paetzold and Lucia Specia. 2016b. Unsupervised Lexical Simplification for Non-Native Speakers. In *AAAI*, pages 3761–3767.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab.

Karl Pearson. 1920. Notes on the History of Correlation. *Biometrika*, 13(1):25–45.

Olivier Picard, Mélanie Lord, Alexandre Blondin-Massé, Odile Marcotte, Marcos Lopes, and Stevan Harnad. 2013. Hidden Structure and Function in the Lexicon. *arXiv preprint arXiv:1308.2428*.

Pranab Kumar Sen. 1968. Estimates of the Regression Coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324):1379–1389.

Charles Spearman. 1906. 'Footrule' for Measuring Correlation. *British Journal of Psychology*, 2(1):89–108.

Sowmya Vajjala and Detmar Meurers. 2014. Assessing the Relative Reading Level of Sentence Pairs for Text Simplification. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 288–297.

Philippe Vincent-Lamarre, Alexandre Blondin Massé, Marcos Lopes, Mélanie Lord, Odile Marcotte, and Stevan Harnad. 2016. The Latent Structure of Dictionaries. *Topics in Cognitive Science*, 8(3):625–659.