International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.04, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Phrase based Statistical Machine Translation for Indian Languages: A Survey

Mr. Satish Kumar[1], Dr. Mohammed Asger[2]

[1]*Department Of Computer Sciences,Baba Ghulam Shah Badshah University*
Rajouri-185131, (J&K), (India)

[2]Department Of Computer Sciences,*Baba Ghulam Shah Badshah University*
Rajouri-185131, (J&K), (India)

ABSTRACT

In this paper, we have analyzed regarding the performance of Phrase-based Statistical Machine Translation on different Indian languages. We present report of baseline systems on several language pairs. The motivation to present this survey is to explore the development of SMT and linguistic resources of these language pairs, the current approaches are quite desolate due to vast amount of data related to linguistic resources. SMT system approaches are unpredictable on large parallel corpus and such data is necessary to reliably estimate translation probabilities. We presented the performance of baseline systems translating from different Indian languages (Hindi, Urdu and Punjabi) into English with accurate results on some extend, for all the language pairs.

Keywords: Statistical Machine Translation, Natural Language Processing, Phrase-based Translation, Parallel Corpus

I.INTRODUCTION TO MACHINE TRANSLATION

In introduction, a brief background of Machine Translation discusses. This topic also includes an overview of different Machine Translation (MT) approaches and also discussed SMT approach being used to carry out this work. The Indian languages that we selected for this task are also discussed briefly in this section.

Machine Translation defined as an automated system that investigate text from a Source Language and applies computation on that input and then yield out text in a required target language without any kind of human intervention. Translation of source text to target

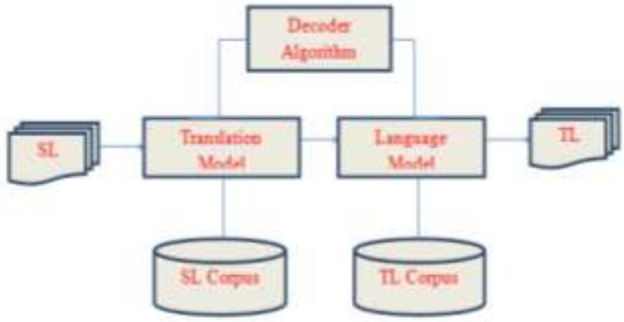

Figure 1:An Architecture of SMT System

language is one of the most interesting and the hardest problem in Natural Language Processing. The adequacy and fluency are two challenges in machine translation. The former challenge reflects to develop a system that adequately represents the ideas expressed in the source language into the target language. The latter challenge reflects to represent those ideas grammatically. Machine translation common approaches are the corpus-based approach and rule-based approach.

Corpus approach, large parallel and monolingual corpora are used as source of knowledge. Corpus approach can be further categorized into statistical approach and example-based approach. In statistical approaches the target text is generated and evaluated by statistical model and parameters of statistical approach are learned through parallel corpus. In Corpus approach, MT is also relate through decision problem, a better target language phrase is decided from the given source language. Statistical decision theory and Bayes rule are applied to solve this decision problem

In the rule-based approach, the text of the source language is analyzed using various tools such as a morphological-parser and analyzer and then transformed into an intermediate representation. This intermediate representation, generate the text in the target language by using some rules. A set of rules is necessary to encapsulate the phenomena of natural language. By applying these rules the grammatical structure of the source language transfer into the target language. As the number of rules grows, the system becomes more complicated and the performance of translation becomes skewed.

II.PHRASE-BASED MODEL

In this section we explain the basic principles of phrase-based models and how they are trained. In machine translation the best performing statistical machine translation systems are based on phrase-based models: - models that translate small word sequences at a time. Phrase-based translation models allow lexical entries with more than one word on either the source-language or target-language side: - for example, we might have a lexical entry

(यह आदमी; this man)

Mention that the string *यह आदमी* in Hindi can be translated as *this man* in English. The option of having multi-word expressions on either the source or target-language side is a significant departure from IBM models 1 and 2, which are essentially word-to-word translation models. Multi-word expressions are extremely useful in translation and this is the main reason for the improvements that phrase-based translation models give.

Phrase-based Lexicon: Formal definition, A phrase based lexicon L is a set of lexical entries, where each lexical entry is a tuple (f, e, g) where

- *f = a sequence of one or more foreign words.*
- *e= a sequence of one or more English words.*
- *g = a "score" for the lexical entry. The score could be any value in the reals.*

Learning Phrasal Lexicons from Translation Examples: Next section, we describe how a phrasal lexicon can be learned from a set of example translations. We'll assume that our training data consists of English sentences $e^{(k)} = e_1^{(k)}\ldots\ldots\ldots e_{lk}^{(k)}$ paired with Hindi sentences $h^{(k)} = h_1^{(k)}\ldots\ldots\ldots\ldots h_{mk}^{(k)}$, for k = 1 ………..n.

Here integer l_k= the length of thek^{th} English sentence,

$_{hlk}^{(k)}$ = the j^{th} word in the k^{th} English sentence

m_k= the length of the k^{th} Hindi sentence

$e_i^{(k)}$ is the ith word in the k^{th} Hindi sentence.

In addition to the sentences themselves, we will also assume that we have an alignment matrix for each training example. The alignment matrix $A^{(k)}$ for the k^{th} example has $l_k \times m_k$ entries, where

$A_{I,j}^{(k)}$= 1 if Hindi word i is aligned to English word j, 0 otherwise

With an alignment matrix $A_{I,j}^{(k)}$, the alignments can be many-to-many; for example, a given Hindi word could be aligned to more than one English word. Figure 1 shows a simple algorithm for this purpose. The input to the algorithm is a set of translation examples, with an alignment matrix for each training example. See figure 2 for the definition of the consistent function. Intuitively, the function checks whether any alignments from English or foreign words in the proposed lexical entry are to words that are "outside" the proposed entry. If any alignments are to outside words, then the proposed entry is not consistent. In addition, the function checks that there is at least one word in the English phrase aligned to some word in the foreign phrase

Inputs: $e^{(k)}$, $h^{(k)}$, $A^{(k)}$ **for k = 1 . . . n**

Initialization: L = ∅

Algorithm:

• For k = 1 . . . n

For s = 1 . . . mk, for t = s . . . m^k

For s = 1 . . . l_k**, for t = s' . . .** l_k

If consistent(A(k) ,(s, t),(s', t')) = True

(1) Define h = h(k) s . . . h(k) t , define e = e (k) s $_0$ **. . . e (k) t** $_0$

(2) Set L = L ∪ {(f, e)}

(3) c(e, h) = c(e, h) + 1 (4) c(e) = c(e) + 1

• For each (h, e) ∈ L create a lexical entry (h, e, g)

where g =

Figure 2: An algorithm for deriving a phrasal lexicon from a set of training examples with alignments.

Phrase-based translation models: A phrase-based translation model is a tuple (L, h, d, ŋ), where:

L is a phrase-based lexicon. Each member of L is a tuple (h, e, g) where h is a sequence of one or more foreign-language words, e is a sequence of one or more English words, and g 2 R is a score for the pair (h, e).

- h is a trigram language model: that is, for any English string e_1 ………...e_m,

$$h(e_1 \ldots\ldots\ldots e_m = \sum_{i=1}^{m} q\,(e_i|e_{i-2}), e_{i-2}$$

Where q are the parameters of the model, and we assume that $e_{-1} = e_0 = *$, where * is a special start symbol in the language model.

- d is a non-negative integer, specifying the distortion limit under the model.

  ŋ∈ R is the distortion penalty in the model.

For an input sentence x1, Y(x) to be the set of valid derivations under the model (L, h, d). The decoding problem is to find

$$\underset{e\epsilon y(x)}{\arg max} f(y)$$

Where, assuming y=p_1 p_2……p_L,

$$f(y) = h(e(y)) + \sum_{k=1}^{L} g(p_k) + \sum_{k=1}^{L} \eta * |t(p_k) + 1 - s(p_{k-1})$$

III.INDIAN LANGUAGE

In this survey we conduct three commonly spoken languages in the sub-continent, parallel corpus of which was available to test.

1.1 Hindi

It is the national and official language of India. 425 million folks speak Hindi as their maternal language whereas quite twelve million folks as their second (Britannica, 2014). Outside India, some communities in African nation, Mauritius, Bangladesh, Yemen, and African country additionally communicate in Hindi language. Hindi is a-member of the-Indo-Aryan-group inside the Indo-Iranian-branch of the-Indo-European-language family. Like in Persian, Hindi adjectives don't modification as results of variety modification in noun. Its preposition is comparable to English. not like others Indic based mostly languages like Guajarati it's solely 2 genders i.e. masculine and female. Case marking in Hindi is straightforward attributable to Persian influence and reduces it to direct kind and an oblique kind. Case relations square measure shown post positions. Like several languages it's additionally written from left to right however its literary genre is SOV. Fashion able commonplace Hindi evolved from the interaction of Muslim from Asian nation, Iran, Turkey, Central Asia, et al. attributable to Persian influence Hindi borrowed some a part of vocabulary from Persian language like dresses. A sizable amount of adjectives and their nominal derivatives, and a good vary of different things and ideas square measure most a locality of the Hindi language that purists of the post-independence amount are unsuccessful in purging them. Where borrowing Persian and Arabic words, Hindi additionally borrowed-phonemes. For example, Hindi renders the word for force as either-zor-or-jor-and the word for sight as-nazaror-najar. In most cases the sounds /g/ and /x/ were replaced by /k/ and /kh/ severally. Contact with a people language has additionally enriched Hindi. Several English words, such as-button, pencil, petrol, and college square measure fully assimilated within the Hindi lexicon.

1.2 Urdu

Urdu is additionally a member of the Indo-Aryan cluster among the Indo-European family of languages. Urdu is the national language of Asian nation whereas it's formally recognized language in Indian constitution similarly. More than one hundred million individuals (Britannica, 2014) among Asian nation and Asian country speak in Urdu. Apart from these 2 nations Urdu is additionally spoken by the immigrants and in little societies in Great Britain, USA and UAE. Urdu and Hindi are bilaterally sonic. This language developed and stemmed from Indian landmass

Therefore it's the same as Hindi. as a result of similarity in teaching reading and synchronic linguistics they appear like one language but there sources are totally different. Urdu is season from Arabic and Persian whereas Hindi is borrowed from Sanskrit that is why they're treated as maverick languages. There’s a large distinction in their genre. Urdu script may be an altered and revised style of Perso-Arabic scripts whereas Hindi script is a changed kind of Deva-nagari script. Urdu and Hindi sound similar except few variations in brief vowel allophones. Urdu with-holds a full set of aspirated stops. It's the property of each Indo Aryan similarly as retroflex stops. Urdu doesn't retain the entire vary of Perso-Arabic consonants, despite its serious borrowing from that tradition. The biggest range of sounds maintained is among the spirants; a gaggle of sounds expressed witha friction of breath against some a part of the oral passage, during this case /f/, /z/, /zh/, /x/, and /g/. One sound in the stops class, the vocal organ /q/, has conjointly been maintained from Perso-Arabic. Grammatically Hindi and

1.3 Punjabi

It is a member of the-Indo-Aryan-subdivision of the-Indo-European language-family. More than 10 million people speak this language (Britannica, 2014) in the domain that was discordant between Pakistan and India during cleave. This language is officially added in Indian constitution. Some small societies in UAE, UK, USA, Canada, South Africa and Malaysia speak Punjabi. It is of two miscellanies; one is western which is known as Lahnda and second is eastern known as Gurmukhi. Thereare two ways to write Punjabi, one is by Perso-Arabic script and other is by Gurmukhi alphabets which were conceived by Sikh Guru Angad (1539-52) rules for scriptural use. Its writing style is SOV and written from left to right (Gurmukhi) and right to left (Perso-Arabic).

IV.DATASET (TEXT COLLECTION) AND EXPERIMENT

For this statistical machine translation approach, parallel corpora were collected from diverse domains for all the selected languages. During the bilingual corpus collection our first motive was to collect data from diverse domains to get better translation quality and a wide range vocabulary. For this purpose the corpus we selected to use in this approach is Enabling Minority Language Engineering (EMILLE). EMILLE contain near about 70 million word corpus of Indic languages which is distributed by the European Language Resources Association. EMILLE contains data from six different categories:- consumer, health, housing, education, social and legal documents. This text collection is based on the information provided by the various local authorities and UK government. In this work we applied 70 parallel files in total for each of our source language and each filename consisting of language code, text type, genre and subcategory, connected with hyphen character. The data is encoded in 2byte Unicode format and marked up in SGML format. Our work based EMILLE corpus that is becoming a standard data repository for languages of Indic region. The parallel corpus includes near approximately words in English and it are accompanying translations in Hindi, Bengali, Punjabi and Urdu. Its bilingual data resources include approximately 14000 sentences for all the available languages from which we were able to sentence-aligned and extract over 8000 sentence for all languages pairing with English using the sentence alignment algorithm. Some experiments with Multi-Indic parallel corpus were also done.

IJARSE
ISSN: 2319-8354

<table>
<tr><th rowspan="3">Source Language</th><th colspan="6">Target Language (English)</th></tr>
<tr><th colspan="2">Training Size (Tokens)</th><th colspan="2">Test Size(Tokens)</th><th colspan="2">Total Sentence Pairs (Tokens) including tuning Sentence tokens.</th></tr>
<tr><th>Source</th><th>Target</th><th>Source</th><th>Target</th><th>Source</th><th>Target</th></tr>
<tr><td>Hindi</td><td>137,623</td><td>102,754</td><td>15,583</td><td>11,517</td><td>172,352</td><td>128,741</td></tr>
<tr><td>Urdu</td><td>124,755</td><td>86,563</td><td>13,465</td><td>9,222</td><td>138,220</td><td>95785</td></tr>
<tr><td>Punjabi</td><td>110,014</td><td>89,136</td><td>13,602</td><td>10,554</td><td>123,616</td><td>99690</td></tr>
</table>

Table 1: Training and Evaluation data for EMILLE

In Statistical Machine Translation development project development of parallel corpus is the most complicated task. The EMILLE corpus that we in this work is quite noisy. Filtering and cleaning of dataset is requires before use. The first step in development of any SMT system is cleaning of the corpus to extract aligned parallel sentences pair. The details about number of parallel sentences that were extracted for each pair are given in Table 1 and Table 2.

Corpus	**Total Sentence**	**Tuning Sentence**	**Tuning Sentence**	**Testing Sentence**
Hindi	9540	7609	952	952
Urdu	8245	6772	847	846
Bengali	8521	6820	860	860
Punjabi	8466	6777	850	850

Table 2: Training and Evaluation data for EMILLE

A sufficiently large English language monolingual corpus is collected for this task. This monolingual corpus apply build the language model that is used by the decoder to select the most affluent translation from several possible translation options. In this work we gather sufficiently large monolingual data from as many different available online resources as possible such as **Europarl**. After this step, train the language model on the corpus that is suitable to the domain. To accomplish this data from diverse domains is collected. The main fields to collect data are Literature, Science, News, Religion, Health and Education. The WMT 08 News Commentary data source used as the main entity for monolingual data, the target side of the parallel corpora is also added to the monolingual data. The monolingual corpora text collection used for this study has approximately 60 million tokens distributed in nearly 2 million sentences.

V.EXPERIMENTAL PROCESS

We carry out k-fold cross validation method for sampling of the corpus for all language pairs. Here k=5 was selected by assign 4/5 as training and 1/5 as tuning and test set for experiment on all folds. Each fold contain over 800 segments for tuning and testing along with above 6500 segments for training for all source languages except Hindi. For Hindi we select more than 9000 segments in total. Nearly 7000 selected for training and about 950 sentences for tuning and testing of Hindi to English translation system.

All these statistics depicted clearly in Table 1. In our work, first step is sampling of data next training, tuning and test sets are tokenized for all folds. At the end, all datasets are converted to lowercase. These steps are repeated for all language pairs using scripts provided by Moses decoder. Finally, lowercase training dataset is used for word alignment.

Baseline Status: In this work we trained a **Moses system** with the following parameters: Sentence length maximum up to 80, GDFA summarization of GIZA++ alignments, an interpolated Kneser-Ney smoothed 5-gram language model with SRILM used at runtime, a 5-gram OSM, msd-bidirectional-fe lexicalized reordering, sparse lexical and domain features, a distortion limit of 6, 100-best translation options, MBR decoding, Cube Pruning with a stack-size of 1000 during tuning and 5000 during test, and the no-reordering-over punctuation heuristic. We tuned with the k-best batch MIRA algorithm. Language Model is built on the available monolingual English corpus and it is implemented as an n-gram model using the SRILM toolkit. For all the experiments in all languages, the same language model is used for all folds of the source languages.

Evaluation Results of SMT System

As the languages used in this work are sparse-resourced, we achieved relatively lower scores for BLEU, we have achieved BLEU score with a mean of 0.12 and a Standard deviation of 0.06 on the given test sets using the 5-fold cross validation method. Table 3 presents the results of experiments for all language pairs. The results are composed of BLEU and NIST score evaluated over the test corpora and also the UNK Count over that test corpus for all the selected language pairs. The subsequent subsections presented evaluation results for all language pairs for both seen i.e. data taken from the training set and the unseen i.e. actual testing data.

Language Pair	**BLEU**		**NIST**		**UNK Count**	
	Mean X	**σ**	**Mean X**	**σ**	**Mean X**	**σ**
Hindi-English	0.115	0.068	3.779	0.804	224	30
Urdu-English	0.140	0.038	4.260	0.535	183	15
Punjabi-English	0.150	0.09	4.185	1.158	197	36

Table 3: Evaluation Results of developed SMT system for all language pairs

Hindi-English:

The results of Hindi to English translation are given in Table 4. The corpora used for Hindi-English language pair was the most domain-relevant and the biggest in size. It resulted in significantly better translation as compare to other language pairs. Hence, it can be concluded that the size and relevance of parallel language corpus have a direct relationship with the quality of translation. For this pair, again wegot decent BLEU scores with a mean of X= 0.115 and a Standard deviation _ = 0.068 on unseen data and X= 0.352 and a Standard deviation _ = 0.025 on seen data. For NIST we got, X= 3.779 and a Standard deviation _ = 0.804 on unseen data and X= 7.634 and a Standard deviation _ = 0.437 on seen data. When counting the unknown words in translation of our SMT system we achieved X= 672 and a Standard deviation _ = 90 on unseen data and X= 150 and a Standard deviation _ = 10 on seen data. Translation output of our developed system is given below in example.

Example:

Source: संपर्कपतेऔरपतेनंबरनिम्नानुसारहैं

Reference: contact addresses and telephone numbers are as follows:

Output: on |0-0| the |2-3| संपर्क |1-1| for |4-5| addresses |6-6| and |7-7| telephone |8-8| helpline |9-9| below |10-10| निम्नानुसार |11-11|

S.No	Input Phrase	Reference Phrase
1	संपर्क	Contact
2	और	And
3	पते	Addresses
4	हैं	Are
5	निम्नानुसार	Follows
6	नंबर	Number

Table 4: Hindi-English Phrase table for given example

Table 4 presents input phrases along with corresponding reference phrases for the example mentioned above. A clear difference can be observed between the reference translation and the one achieved from the developed system. The translation output is segmented into different phrases and decoder fetches the translation from the developed phrase table. The reordering model also gave poor result for such small amount of data.

In output the first word of source is translated to “on” then next two words were translated as "the" then again NULL token so it becomes an OOV in our translation output. From the phrase table it is seen that many source words are translated to just single target output. This is also because of poor tokenization for regional languages as there is no standardized tokenizer available for these languages. Table 10 shows the actual BLEU, NIST score for all the folds along with the OOV words count.

Folds	**BLEU**		**NIST**		**UNK Count**	
	Seen	**Unseen**	**Seen**	**Unseen**	**Seen**	**Unseen**
F1	0.360	0.065	7.765	3.528	134	782
F2	0.363	0.055	8.760	3.528	155	764
F3	0.354	0.064	7.885	3.528	144	734
F4	0.335	0.068	7.435	5.528	163	644
F5	0.329	0.025	7.365	5.528	123	587

Table 5: Evaluation results for Hindi-English translation

Urdu-English

For Urdu-English language pair, we plotted BLEU scores with a mean of X= 0.15 and Standard deviation **σ**= 0.040on unseen data then we got X= 0.381 and a Standard deviation **σ**= 0.027 on seen data. For NIST we got, X= 4.36and a Standard deviation **σ**= 0.545 on unseen data and got X= 7.54 and a Standard deviation **σ**= 0.53 on seen data with small amount of training parallel corpus. Using our statistical machine translation we counting the unknown words and we come up with X= 560 and a Standard deviation **σ**= 44 on unseen data and got X= 117 and a Standard deviation **σ**= 12 on seen data. The following example shows the different type of problems that are come up in translation output from the developed system.

Example:Source: .20. بہتری کی یہ باتیں ایک عمدہ ابتدا ہیں ۔

Reference: 20. These improvements are a good start.

Output: 20. |0-0| the |1-2| these |3-3| things to |4-4| start |7-7| a |5-5| quality |6-6|. |8-9|

S.No	Input Phrase	Reference Phrase
1	20.	Null
2	بہتری کی	the need for improvements
3	یہ	These
4	باتیں	Things to
5	ایک	A
6	عمدہ	Good quality
7	ابتدا	Start
8	ہیں ۔	.

Table 6: Urdu-English Phrase table for given example

In output the first word of source and target is same so decoder did nothing with it and its segment from phrase table will be NULL. The next word got totally different output in translation output as compared to the phrase table entry of Table 6. The two source words are translated to four word phrase in phrase table but in our translated output we got just a single output translation. This is because of the n-best translation phrase for a single phrase input. Next we can see the reordering again poorly managed by the baseline phrase based model. All this discussion with given output example lead us to a bottom line conclusion that if we manage to get a good tokenizer and more corpora for all the selected regional languages, it will lead us to decent BLEU scores and fluent translations. Table 7 shows the actual BLEU, NIST score for all the folds along with the OOV word count.

Folds	**BLEU**		**NIST**		**UNK Count**	
	Seen	**Unseen**	**Seen**	**Unseen**	**Seen**	**Unseen**
F1	0.360	0.065	7.765	3.528	134	782
F2	0.363	0.055	8.760	3.528	155	764
F3	0.354	0.064	7.885	3.528	144	734
F4	0.335	0.068	7.435	5.528	163	644
F5	0.329	0.025	7.365	5.528	123	587

Table 7: Evaluation results for Urdu-English translation

VI.FUTURE WORK AND CONCLUSION

We have shown how developed statistical machine translation system takes the Indian language sentences as input and it generates corresponding closest translation in English. Using automatic evaluation translation of more than 800 sentences was evaluated. The average evaluation (using BLEU) score of 15% to 20% was reported for all the Indian languages. This low evaluation scores show that the quality of translation is directly dependent on the scope and quality of parallel language corpora.

All the Indian Languages we have used in this work exhibit rich morphology thus resulting in sparse estimates which causes poor translation quality. Therefore, results are not as good as the reported in the Euro-Languages for which parallel and monolingual data available. In this paper, we carried out a set of experiments by choosing the training, tuning and test sets from parallel corpus applying fivefold cross-validation method.

Results of SMT shown that each of our sources Indian language got most divergence once translating into English which is why there's a big distinction in obtained MT analysis scores on the seen corpus and on unseen take a look at sets.

In future, we will explore and study different statistical machine translation approach by applying other different approaches to develop better language models and also the training model for all Indian languages whose parallel corpus is available in nearer future. We also intend to perform a good evaluation technique for better scoring analysis on the MT output for all the language pairs we used for experimenting to compare our MT evaluation results for both the seen and unseen datasets as there are unknown words occurring in a translation of seen test sets.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[8] Electronic Publication: Digital Object Identifiers (DOIs):

[9] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[10] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.