International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

LANGUAGE SEGMENTATION: WORDS WITH AVERAGE PREDICTED TIME STAMP PERIOD

DR.T. PREM CHANDER
Associate Professor, ISL Engineering College, Hyderabad, India
Email: tudiprem@gmail.com

Abstract: NLP (Natural Language Processing) is a technology used by computers to understand human language. Words are the basic unit for Language segmentation. Structural difference of data in various environments may not map directly, this paper introduces the word learning and getting multi word from various sources. Each word with multi sub structure converting to same word pattern. Most common used words are made as long-term word set and language segmentation processing is proposed. Average prediction processing method on word segmentation for processing speed is shown efficient.
Index Terms: **language segmentation, long-tem word set, prediction processing method.**

I. INTRODUCTION

With the develeopmen of IOT and technology, a large amount of data types and strucutes have been accumulated in scientific areas. The observation of same semantic is difficult, various methods often used to get data from multiple readings, and these data has different channels with the same concepts. Each data set said as model and perspectives. Different modals together have multi-modal data for same specifications. Natural Language Processing is a technology for human-computer interaction and AI.

NLP has focus on the analysis of language recognition and social media data. The training of each word learning with effectiveness is more important. Another problem is connection between more layer structure with representations.

The methods available for word segmentation algorithm based on supervised learning are as follows:

1. The method with highest recall rate of 72.9% closes test extraction with algorithm an dictionary structure.
2. English word segmentation by rule-based methods suitable dictionaries according to special scenarios.
3. Big data as natural processing for word segmentation by parallel computing and able to perform parallel learning on large-scale training data.
4. Global information as dictionary data for word matching in statistical and structural sentences.
5. Labeling through indexing algorithms based on fragments in query processing systems.

Most of the word segmentation methods has less dependency of text with time sequences. We propose a model to improve the processing efficiency of word segmentation. Since machine learning models cannot have dependency on time sequencing in both directions. We use a method for recurring words and combine the occurrences count of words in different models.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

II. WORD EXTRATION BY NLP

A. *Natural Language Processing*

NLP has Text segmentation and word segmentation The Text segmentation faces problem of word segmentation. Segmentation divides a given sentence to meaning word phrases. Text alignment is very important with the processed scope to the content of words stored. How to get the correct segmentation is the most prior affects of the word segmentation system. When we process text in any language grammar is important and words are basic unit. Words are the main concept of information retrieval. The text retrieval and extraction of words are basic terms.

The basic word segmentation methods are by words, scanning from right to left, scanning from left to right, comparisons.

Drawback for word segmentation is cost and time. Tagging the words with grammar and identifying the vocabulary type. Collecting the taggers count and finding probability for each grammar terms.

Possible Features in Word are as follows:

1. Binary
2. Categorical or continuous
3. Structured data with size and forms
4. Classifiable and quantitative

Considering only the words for segmentation not taking other set of data for processing is very important. Sentences or data sets with more features are learned using flow pattern embedding learning as most effective. Word extraction has pattern learning with word metric mapping and word counting sequence.

Word metric takes various word patterns and results in pictorial notation. Word pattern, word length and scanning process are input for word metric. This method results in two word samples which are occurring in most. Less word length data sets will be more efficient for this method.

Drawback for this process is only takes less length words and computation with storage is huge.

Word counting sequence has sample pair or words with similar length by scanning. If this is done then word count is incremented by one. Drawback of this method is words with different characteristics and means exactly same.

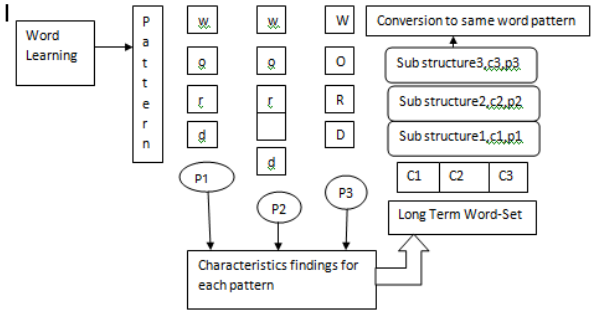

Figure1. Word Segmentation Process with Sub Layers

As shown in the Figure 1, the word segmentation proposed with the word learning and getting multi word from various sources. Each word with multi sub structure converting to same word pattern. Most common used words are made as long-term word set and language segmentation processing is done.

The following are the phases in each word Learning process:

1. Pattern collection for each word sub layer.
2. Characteristic findings for each pattern.
3. Long Term Word-Set formation.
4. Counting sequence for each word.
5. Sub structuring as:
   - 5.1. Sequence of substructure.
   - 5.2. Count of word.
   - 5.3. Pattern matching.
6. Conversion to same word patterns.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

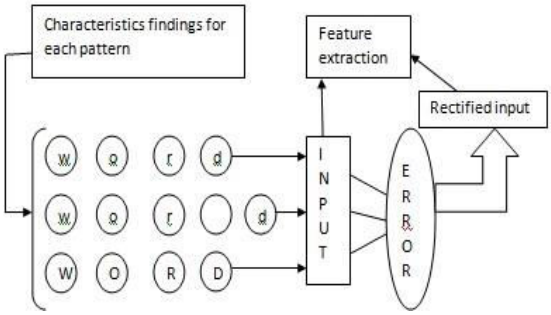

Figure2. Characteristics findings for each pattern

We propose model for different raw data words with feature extraction and eliminate the differences in each pattern.

A characteristic finding has following steps:

1. Input taken as set of words from various data sets in the form of sentences.
2. Finding the spacings between character by distance vector algorithms.
3. Collecting right to left and left to right scanning values in each word set.
4. Counting the occurrence of each word in different patterns.
5. Updating the pattern count.
6. Optimization of feature with Long word set.
7. Finding error in each word pattern
8. Rectifying the error (spaces, length and word count)
9. Input to the feature extraction process.
10. Word set formation.

III. Result Analysis

The ability to find the word segmentation in various techniques is compared to the proposed method. The main advantage is the sub structuring by different algorithms and applied data set are accurate.Already existing methods NUS-WIDE-Object has word modularity and image modularity for text data. The data loss in PCA method and data classification using PCA+SCM is less accurate for image and word segmentation. The characteristics or feature extraction in this model is very accurate.

The dimensions of each model are compared in terms of accuracy and data sets are as follows:

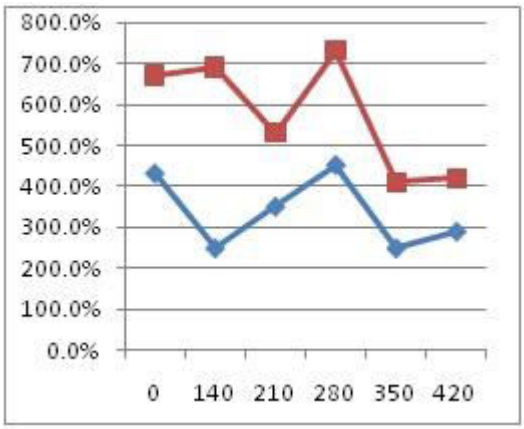

Figure3. Dimensionality and accuracy

The obtained dimensionality and accuracy features is measured with effectively integrated modals. The model is tested by reduce in dimensionality and multi model data.

The proposed model has the performance and validity of each model is shown with each dimensionality. In order to verify of given model in this paper we have set of models with observed values and features. The experimental results indicated with dimensionality and training data given as range values. The classification reduce is the missing data model.

The difference in dimensionality is taken as for further processing sequence and concluded with the values where only one set of data is given. The resultant set of data is the most occurrence count of each word taken in considering grammar of Language.

IV. Word Processing

The segmentation is done based on the characters count in each sequence of sentences.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Each count is taken by the set of defined grammar rules of a language. The commonly used annotation set for sentence, middle of sentence, end of sentences and individual set of words.

English word segmentation is of three parts.

1. Characters converted to vector matrix through words.
2. Conversions of words with multiple words pattern.
3. Words classification by index tagging.

Algorithm of Word segmentation and processing

1. Given a sentence of characters as input.
2. Window with loaded character set and scanning each character from left to right and right to left.
3. Applying the linear conversion from the set if input sentences.
4. Finding the weight matrix and deviation coefficient for each sentence.
5. Getting the character level result by use of index labeling to each input character.

Proposed Model for Word Segmentation:

1. Data set and practical steps in extraction of words from each model.
2. Sentence as input and perform word vector conversion through word2vec, word changed to space vector with length in vector matrix.
3. Obtaining the feature vector for each sub set of words in the collection of sequence framing.
4. Grammar based segmentation on each words segmentation process.
5. Taking random data and framing the selected training set of data models.

V. CONCLUSIONS

This paper proposes a theoretical concept of multi model features extraction from sentences based on given data model and structure of training data model. In order to verify this model as series of comparative experiments has to be carried out. The proposed model with some practical data has been more effective and low-dimensional fusion feature from the original data models.

REFERENCES

[1] Viet Khanh Ha, Jin-Chang Ren, Xin-Ying Xu. “Deep Learning Based Single Image Super-resolution: A Survey", *International Journal of Automation and Computing*, vol.16, no.4, pp.413-426,2019.

[2] Fan Meng, Peng Chen, Lenan Wu. “Automatic Modulation
Classification: A Deep Learning Enabled Approach", *IEEE
Transactions on Vehicular Technology*, vol.67, no.11, pp.10760-
10772, 2018.

[3] Qing Xia, Shuai Li, Aimin Hao. “Deep Learning for Digital
Geometry Processing and Analysis: A Review", *Journal of
Computer Research and Development*, vol.56, no.1, pp.155-182,
2019.

[4] Yuanfang Chen, Yan Zhang, Sabita Maharjan. “Deep Learning for Secure Mobile Edge Computing in Cyber-Physical TransportationSystems", *IEEE Network*, vol.33, no.4, pp.36-41, 2019.

[5] Weber KA, Smith AC, Wasielewski M. “Deep Learning

Convolutional Neural Networks for the Automatic Quantification of Muscle Fat Infiltration Following Whiplash Injury", *Scientific reports*, vol.9, no.1, pp.7973, 2019.

[6] Olivier Bernard, Alain Lalande, Clement Zotti. "Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved?", *IEEE Transactions on Medical Imaging*, vol.37, no.11, pp.2514-2525, 2018.

[7] Murat Abdughani, Jie Ren, Lei Wu. "Supervised Deep Learning in High Energy Phenomenology: a Mini Review", *Communications in Theoretical Physics*, vol.71, no.8, pp.955, 2019.

[8] Rajeev Ranjan, Swami Sankaranarayanan, Ankan Bansal. "Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans", *IEEE Signal Processing Magazine*, vol.35, no.1, pp.66-83, 2018.

[9] Yuling Tian, Xiangyu Liu. "A deep adaptive learning method for rolling bearing fault diagnosis using immunity", *Tsinghua Science and Technology*, vol.24, no.6, pp.750-762, 2019.

[10] Chenming Wu, Rui Zeng, Jia Pan. "Plant Phenotyping by Deep-Learning Based Planner for Multi-Robots", *IEEE Robotics and Automation Letters*, vol.4, no.4, pp.3113-3120, 2019.

[11] Honghao Gao, Wanqiu Huang, Xiaoxian Yang.Applying Probabilistic Model Checking to Path Planning in an Intelligent Transportation System Using Mobility Trajectories and Their Statistical Data. Intelligent Automation And Soft Computing(Autosoft), 2019, 25(3):547–559

[12] Honghao Gao, Wanqiu Huang, Yucong Duan, Xiaoxian Yang, Qiming Zou. Research on Cost-Driven Services Composition in an Uncertain Environment. Journal of Internet Technology(JIT), 2019,20(3):755-769.

[13] Honghao Gao, Yucong Duan, Lixu Shao, Xiaobing Sun.Transformation-based processing of typed resources for multimedia sources in the IoT environment. Wireless Networks, 2019. [online] https://doi.org/10.1007/s11276-019-02200-6

[14] Honghao Gao,Yueshen Xu, Yuyu Yin, Weipeng Zhang,Rui Li,Xinheng Wang. Context-aware QoS Prediction with Neural Collaborative Filtering for Internet-of-Things Services. IEEE Internet of Things Journal, 2019. [online] https://doi.org/10.1109/JIOT.2019.2956827

[15] Xaojin Ma, Honghao Gao, Huahu Xu, Minjie Bian. An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing. EURASIP

Journal on Wireless Communications and Networking, 2019.
[online] https://doi.org/10.1186/s13638-019-1557-3

[16] Siqi Nie, Meng Zheng, Qiang Ji. "The Deep Regression Bayesian Network and Its Applications: Probabilistic Deep Learning for Computer Vision", *IEEE Signal Processing Magazine*, vol.35, no.1, pp.101-111, 2018.

[17] Ashok Kumar, Saurabh Mukherjee, Ashish Kr. Luhach. "Deep learning with perspective modeling for early detection of malignancy in mammograms", *Journal of Discrete Mathematical Sciences and Cryptography*, vol.22, no.4, pp.627-643, 2019.

[18] Zengqiang Yan, Xin Yang, Kwang-Ting Cheng. "Joint Segment-Level and Pixel-Wise Losses for Deep Learning Based Retinal Vessel Segmentation", *IEEE Transactions on Biomedical Engineering*, vol.65, no.9, pp.1912-1923, 2018.

[19] Maximilian Treder, Jost Lennart Lauermann, Maged Alnawaiseh. "Using Deep Learning in Automated Detection of Graft Detachment in Descemet Membrane Endothelial Keratoplasty: A Pilot Study", *Cornea*, vol.38, no.2, pp.1, 2018.

[20] Pei Zhou, Xuming Fang, Xianbin Wang. "Deep Learning Based Beam Management and Interference Coordination in Dense mmWave Networks", *IEEE Transactions on Vehicular Technology*, vol.68, no.1, pp.592-603, 2018.

[21] Jun Yu, Chaoyang Zhu, Jian Zhang, Qingming Huang, and Dacheng Tao, Spatial Pyramid-Enhanced NetVLAD with and Weighted Triplet Loss for Place Recognition, IEEE Transactions on Neural Networks and Learning Systems, DOI: 10.1109/TNNLS.2019.2908982, 2019.

[22] Jun Yu, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, Jianping Fan, Leveraging Content Sensitiveness and User Trustworthiness to Recommend Fine-Grained Privacy Settings for Social Image Sharing, IEEE Transactions on Information Forensics and Security, vol. 13, no. 5, pp. 1317 – 1332, 2018.

[23] Jun Yu, Chaoqun Hong, Yong Rui, Dacheng Tao, Multitask Autoencoder Model for Recovering Human Poses, IEEE Transactions on Industrial Electronics, vol. 65, no. 6, pp. 5060-5068, 2018.

[24] Masumoto Hiroki, Hitoshi Tabuchi, Shunsuke Nakakura. "Deeplearning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity", *Journal of Glaucoma*, vol.27, no.7, pp.1, 2018.