Sign Language Keyword Extraction based on GLOSS

Ruizhu Wu

Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, 100101, China.

504648339@qq.com

Abstract. In order to quickly understand the content of sign language video, the theme of handshake language, and facilitate the efficient management and retrieval of sign language corpus, the annotation corpus in the parallel corpus of the text first maps all words to one using the word2vec model based on deep learning tools. Abstract word vector space; then word clustering based on K-means algorithm to achieve keyword extraction. Experiments show that the algorithm has better keyword extraction effect for sign language videos with more keywords and longer video time.

Keywords: Gloss; sign language; keyword extraction; Word2vec.

1. Introduction

Sign language video is a special form of video. Let's first explain how general video is extracted from keywords. The extraction of video semantic information can be divided into two major categories. One is to obtain semantics or to reflect semantic clues by extracting some very special expression semantic objects in the video. For example, the text in the video is identified and then indexed to support keyword-based retrieval. This type of method generally works better, but only for certain video types. The other is to use the machine learning method to classify the various low-level features of the video based on the semantic concept defined in advance, so that the acquired semantic information can still be index-based indexed through a combination of various concepts. Search. The advantage of this type of method is that the video content itself is analyzed and can be applied to all types of video, but requires some manual intervention.

The word extraction using Word2Vec and K-means algorithm in this paper is different from the above two methods. It is the extraction of semantic information by using the annotation information of sign language video. The annotation information of the sign language video is performed when the sign language corpus is established, and the sign language information is recorded to facilitate subsequent research. We use the Gloss layer opponent language in the transfer layer to extract the semantic information.

2. Research Content

2.1 Introduction to Sign Language Corpus Annotation Tools and Methods

ELAN (EUDICO Linguistic Annotator) [1] was developed by the Max Planck Institute of the Netherlands for the study of psycholinguistics, with the aim of providing good technical support for annotation and development of multimedia. ELAN is designed for analysis of language, sign language, and gestures, but it can be used by research media corpus researchers to easily annotate, analyze, and record in video and audio data. ELAN is a professional tool for complex labeling of video or audio files. Use ELAN to add unlimited layers of annotations to your video and audio. The content of the annotation can be a sentence, a word, a content, a translation or a description of the details of the video, and the like. Using the ELAN opponent video to mark up can achieve twice the result with half the effort [2].

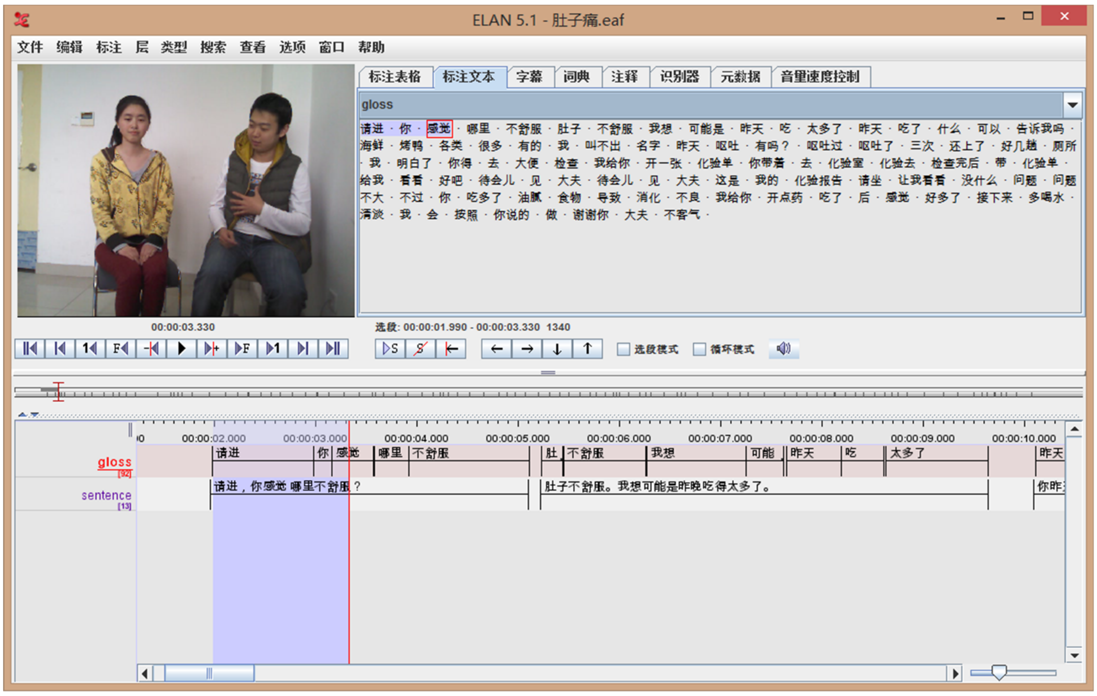
(1) Tier is the basis for transfer and labeling, and different layers have different label contents. Such as annotation layer, word class layer, translation layer and so on. The layers in the ELAN can be added according to the needs of the user.

(2) Transcription refers to the operation of entering text or other symbols according to audio and video. Taking sign language as an example, the content and method of sign language expression are

recorded in the order of sign language by using Chinese characters and other characters. Without translation processing, the original hand sentences are written, not translated Chinese sentences [3].

(3) Annotation is a text, a comment, a translation, an international phonetic transcription, etc., which are transferred for audio or video content, and the annotation includes a transcription. In ELAN, the label also refers to the timeline on the time period, and no content can be transferred during the time period.

We will use the ELAN software to add a label layer to the opponent language video, which is the gloss and the sentence layer respectively. The gloss layer is the meaning of the gesture in the time period, which is based on the national common sign language; the sentence layer is the sentence translation. The gloss layer is exported to a text corpus for further study. As shown in Figure 1, the information in the gloss layer is displayed in the label text.



请进 · 你 · 感觉 · 哪里 · 不舒服 · 肚子 · 不舒服 · 我想 · 可能是 · 昨天 · 吃 · 太多了 · 昨天 · 吃了 · 什么 · 可以 · 告诉我吗 · 海鲜 · 烤鸭 · 各类 · 很多 · 有的 · 我 · 叫不出 · 名字 · 昨天 · 呕吐 · 有吗？ · 呕吐过 · 呕吐了 · 三次 · 还上了 · 好几趟 · 厕所 · 我 · 明白了 · 你得 · 去 · 大便 · 检查 · 我给你 · 开一张 · 化验单 · 你带着 · 去 · 化验室 · 化验去 · 检查完后 · 带 · 化验单 · 给我 · 看看 · 好吧 · 待会儿 · 见 · 大夫 · 待会儿 · 见 · 大夫 · 这是 · 我的 · 化验报告 · 请坐 · 让我看看 · 没什么 · 问题 · 问题不大 · 不过 · 你 · 吃多了 · 油腻 · 食物 · 导致 · 消化 · 不良 · 我给你 · 开点药 · 吃了 · 后 · 感觉 · 好多了 · 接下来 · 多喝水 · 清淡 · 我 · 会 · 按照 · 你说的 · 做 · 谢谢你 · 大夫 · 不客气 ·

选段: 00:00:01.990 - 00:00:03.330 1340

请进，你感觉 哪里不舒服？

肚子不舒服。我想可能是昨晚吃得太多了。

你昨

Figure 1. Example of sign language labeling

2.2 Introduction to the Algorithm

2.2.1 Introduction to Word2vec

With the development of deep learning in recent years, the feature word extraction based on neural network model, that is, the way "word vector" represents text has been paid more and more attention by academic circles. The distributed representation was first proposed by Hinton in 1986 [4]. The basic idea is to map each word into a K-dimensional real number vector by training (K is generally a hyperparameter in the model), and judge the distance between the words by the distance between the words (such as cosine similarity, Euclidean distance, etc.). Semantic similarity, and word2vec uses the word vector representation of this distributed representation [5~7]. The application of word vectors to natural language processing has been very successful and has been widely used in part-of-speech analysis, finding similar words, keyword clustering and classification.

The specific principle of the algorithm is shown in Equation 1.

$$\mathrm{p}(w_t \mid w_t - \frac{c}{2}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_t + \frac{c}{2}) = \prod\nolimits_{i=1}^{c} p(context(w)i) \tag{1}$$

The CBOW consists of three layers: input layer, projection layer and output layer, and forms effective association and nesting with each other. The input layer is the representation of the word vector. The main function of the projection layer is to superimpose the input vector. The formula used is formula (2).

$$\sum_{i=1}^{c}(context(w)i) \tag{2}$$

Where context(w) represents the probability value of the word, and the output layer is represented by a tree structure, specifically a binary tree. In the specific implementation is done by building a Huffman tree. As word2vec is widely used in text clustering, information retrieval and topic mining, the various word2vec versions of the package have been applied in different specific explorations. Our word2vec based on gensim[8] was chosen for the specific exploration of this article.

2.2.2 Introduction to K-means

The K-means algorithm is an algorithm that solves non-convex optimization problems by using an alternating minimization method. It divides a given data set into user-specified k clusters, which has high execution efficiency and high execution efficiency. Historically, researchers in many different fields have basic K-means algorithms. Researches have been carried out, among which are known as Forgey (1965), McQueen (1967), etc. [9]. Jain and Dubes describe in detail the development history and various variants of the K-means algorithm. The algorithm steps are as follows [10] .
Algorithm input: data set X, number of clusters k

Algorithm output: cluster represents set C

Step1: randomly select k sample objects from the entire data set X as the initial cluster center;

Step2: Calculate the distance from each sample object xm in the data set to the cluster center ci;

Step3: Find the minimum distance from each sample object xm to the cluster center ci, and classify the sample object xm into the same cluster as ci;

Step4: Calculate the mean of the objects in the same cluster, and update the cluster center;

Step5: Repeat Step2~Step4 until the cluster center no longer changes.

2.3 Sign Language Keyword Extraction Process

In this paper, using the glossary corpus, Word2Vec's deep learning method is used to extract keywords. All words are projected into the K-dimensional vector space. Each word can be represented by a K-dimensional vector, that is, Word2Vec is a word. The deep learning model transformed into vector form can simplify the processing of corpus text content to calculate the vector operation of word vector space, and calculate the similarity of vector space to represent the semantic similarity between keywords.

The extraction process of sign language video keywords based on Word2Vec algorithm is shown in Figure 2.

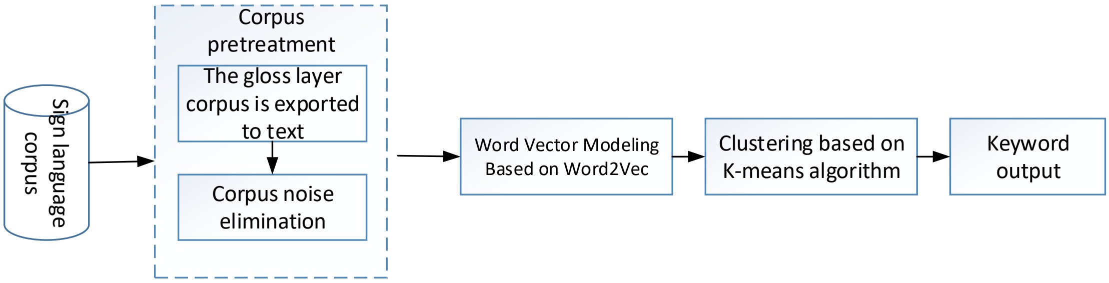

Sign language corpus
Corpus pretreatment
The gloss layer corpus is exported to text
Corpus noise elimination
Word Vector Modeling Based on Word2Vec
Clustering based on K-means algorithm
Keyword output

Figure 2. The extraction process of sign language video keywords based on Word2Vec algorithm

The specific content is as follows: the transfer layer of the corpus to be trained is exported, and when the transfer layer is exported into text, the space is used as a separator, each word has been segmented, or the word segmentation according to the sign language is used to segment the word, and then the corresponding word is removed. Stop words, such as "", "le", "ne", "a" and other auxiliary words. The Word2Vec algorithm model is used to train the opponent language corpus, and the word vector of the sign language is modeled to obtain the word vector corresponding to each word in the corpus. After the word vector is obtained, the similarity calculation is performed. In this paper, the K-means clustering algorithm is used to calculate the cosine distance between the word vectors in the clustering process, and finally the word closest to the clustering center is used as the keyword.

3. Experimental Results and Analysis

The experimental environment is the Anaconda platform, the gensim library provided by the Python language to implement the keyword extraction experiment, and the experiment uses the accuracy as the evaluation standard. In the evaluation of keyword accuracy, this paper takes the number of keywords and the duration of sign language video as the evaluation factor, and manually signs 3, 4, and 5 sign language keywords through sign language users, and the specific results are as follows. Table 1 shows. The result is calculated according to formula (3).

$$\text{Accuracy}=\frac{\text{The same number of keywords as the sign language user}}{\text{Total number of keywords}}*100\% \qquad (3)$$

Table 1. Comparison of keyword accuracy between different keyword numbers and different sign language video durations

Accuracy		The number of keywords		
		3	5	7
The time of Sign language video time	1min	0	6.7	9.5
	3min	11.1	13.3	14.3

It can be seen from the experiment that the algorithm works better when the number of keywords increases and the duration of the sign language video increases. Aiming at this phenomenon, we deeply think and analyze the process of keyword keyword extraction in the opponent language, and get the following conclusions: For the sign language corpus with long sign language video and rich content expression, the accuracy of Word2Vec algorithm based on deep learning model is high; The more the keywords, the higher the accuracy of the Word2Vec algorithm based on the deep learning model.

4. Conclusion and Outlook

This paper first introduces the annotation method based on ELAN software in the opponent language, and then uses the deep learning model Word2Vec word vector model and K-means clustering algorithm to realize the keyword extraction of sign language corpus, which can help researchers quickly understand the sign language video content and grasp The sign language theme improves the efficiency of the corpus retrieval and also helps the corpus to classify. It also provides reliable technical means for everyone to understand and analyze sign language. As a kind of natural language, sign language should attract the attention of academic circles. Sign language research has just started, and the protection of sign language and deaf culture is imminent. There are still many things that we need to do.

References

[1]. Birgit Hellwig ELAN - Linguistic Annotator version 4.5.0[M] 2013-01-07 The latest version can be downloaded from: http://tla.mpi.nl/tools/tlatools/elan/.

[2]. Li Heng, Wu Ling. Basic Methods of Sign Language Corpus Construction[J]. China Special Education, 2013(3): 38-42.

[3]. Gong Qunhu, Yang Junhui. Chinese Sign Language C hinese Transfer Scheme. 2009-2-13.

[4]. Hinton G E.Learning distributed representations of concepts.[C]// Eighth Conference of the Cognitive Science Society. 1989.

[5]. TANG Ming, ZHU Lei, ZOU Xianchun. Document Vector Representation Based on Word2Vec[J]. Computer Science,2016(6):214-217,269.

[6]. LI Yuepeng, JIN Cui, JI Junchuan. A Keyword Extraction Algorithm Based on Word2vec[J]. E-Science Technology& Application,2015(4):54-59.

[7]. ZHENG Wenchao, XU Peng. Research on Chinese word Clustering with Word2vec[J]. Software,2013(12):160-162.

[8]. eh ek R. models.word2vec-Deep learning with word2vec[EB/OL]. [2018-07-26]. https:// radimrehurek. com/ gensim/ models/ word2vec.html.

[9]. Wu X D, Vipin Kumar. Ten Algorithms for Data Mining [M]. Beijing: Tsinghua University Press, 2013.

[10]. Jaroš M, Strakoš P, Karásek T, et al. Implementation of K-means segmentation algorithm on Intel Xeon Phi and GPU: Application in medical imaging[J]. Advances in Engineering Software, 2017, 103(C):21-28.