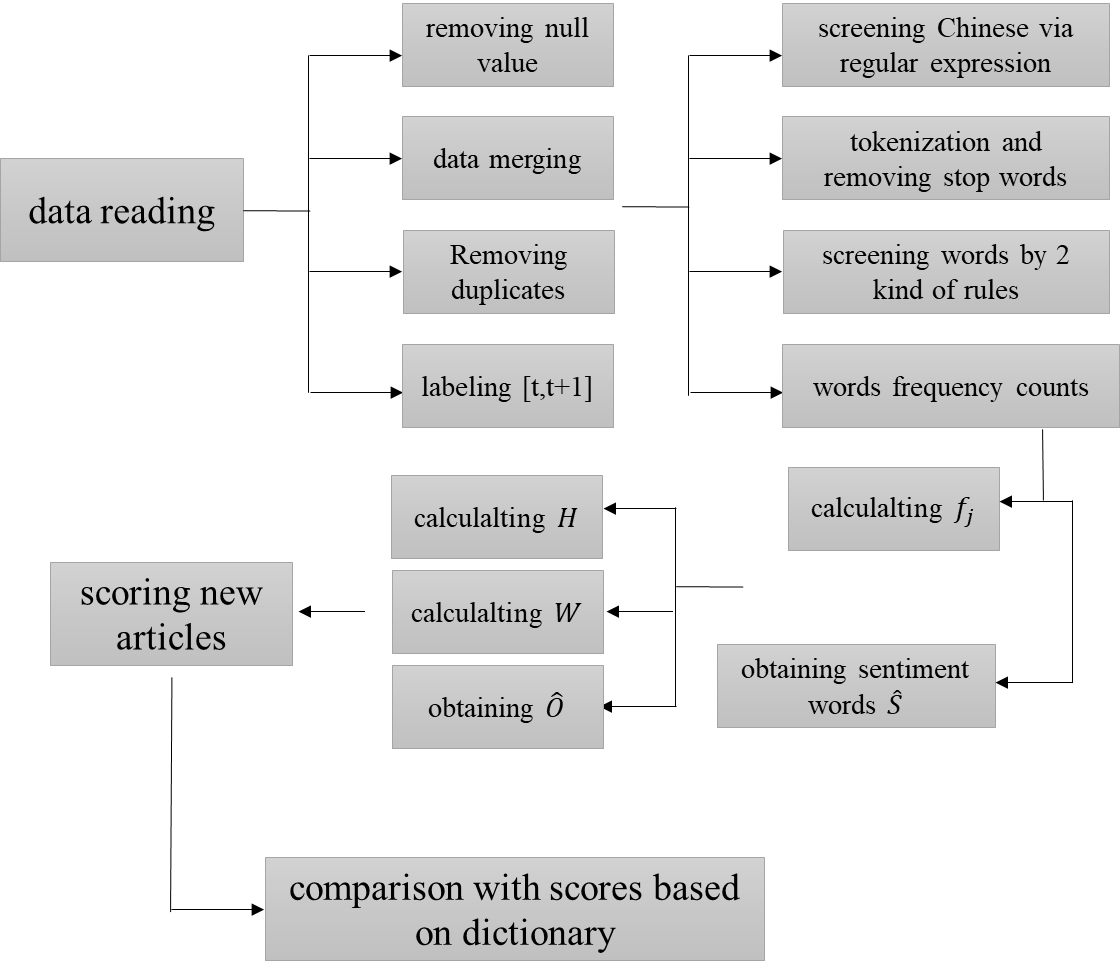
**Research Framework**



**Algorithm: SESTM**

Step 1: Screening for Sentiment-Charged Words

Step2: Learning Sentiment Topics

Step3: Scoring New Articles

**Empirical Analysis**

1. Using 2 rules to screen words

* Words with [a+v]
* Based on word length

1. Using specret or specret [t,t+1] (delayed effect) as label data
2. Trying different number of sentiment charged words including 100 positive 100 negative or 500 positive 500 negative
3. Trying different quantile of word distribution including 94%, 92%, 90%, 88%, 86% to get sentiment charged words
4. Using 2 sentiment word dictionaries as benchmark for comparison

The partial results as below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N=100 | N=500 | Benchmark1 | Benchmark2 | Benchmark3 |
| Spearman | 0.101693 | 0.099348 | 0.027894 | -0.118179 | -0.034445 |
| Pearson | 0.093179 | 0.102048 | 0.038037 | -0.119394 | -0.018576 |

In this experiment,

N is the number of positive words or negative words;

Benchmark1-3 are the score1-3 in your email;

Using specret [t,t+1] (delayed effect) as label data;

Using 88% quantile of word distribution;

Using the positive or negative words in annual report as dictionary.

Obviously, the results of SESTM are better than the results based on dictionary.