

Then we compute the candidate confidence by using the following iterative formula.

$$C^{t+1} = M^T \times M \times C^t \quad (6)$$

where C^t is the candidate confidence vector at time t , and C^{t+1} is the candidate confidence vector at time $t+1$. M is an opinion relevance matrix, a $m \times n$ matrix, where $M_{i,j}$ is the associated weight between a noun/noun phrase i and an adjective j .

To consider the candidate importance scores, we introduce a reallocate condition: combining the candidate opinion relevance with the candidate importance at each step. Thus we can get the final recursive form of the candidate confidence as follows.

$$C^{t+1} = (1-\lambda) \times M^T \times M \times C^t + \lambda \times S \quad (7)$$

where $\lambda \in [0,1]$ is the proportion of candidate importance in the candidate confidence. When $\lambda=1$, the candidate confidence is completely determined by the candidate importance; and when $\lambda=0$, the candidate confidence is determined by the candidate opinion relevance. We will discuss its effect in the section of experiments.

To solve Eq. (7), we rewrite it as the following form.

$$C = \lambda \times (I - (1-\lambda) \times M^T \times M)^{-1} \times S \quad (8)$$

where I is an identity matrix. To handle the inverse of the matrix, we expand the Eq. (8) as a power series as following.

$$C = \lambda \times [I + B + \dots + B^k] \times S \quad (9)$$

where $B = (1-\lambda) \times M^T \times M$ and $k \in [0, \infty)$ is an approximate factor. In experiments, we set $k=100$. Using this equation, we estimate confidences for opinion target candidates. The candidates with higher confidence scores than the threshold will be extracted as the opinion targets.

4 Experiments

4.1 Datasets and Evaluation Metrics

In our experiments, we select three real world datasets to evaluate our approach. The first dataset is *COAE2008 dataset2*², which contains Chinese reviews of four different products. The detailed

information can be seen in Table 2. Moreover, to evaluate our method comprehensively, we collect a larger collection named by *Large*, which includes three corpora from three different domains and different languages. The detailed statistical information of this dataset is also shown in Table 2. Restaurant is crawled from the Chinese Web site: www.dianping.com. The Hotel and MP3³ were used in (Wang et al., 2011), which are respectively clawed from www.tripadvisor.com and www.amazon.com. For each collection, we perform random sampling to generate testing dataset, which include 6,000 sentences for each domain. Then the opinion targets in *Large* were manually annotated as the gold standard for evaluations. Three annotators are involved in the annotation process as follows. First, every noun/noun phrase and its contexts in review sentences are extracted. Then two annotators were required to judge whether every noun/noun phrase is opinion target or not. If a conflict happens, a third annotator will make judgment for final results. The inter-agreement was 0.72. In total, we respectively obtain 1,112, 1,241 and 1,850 opinion targets in Hotel, MP3 and Restaurant. The third dataset is Customer Review Datasets⁴ (English reviews of five products), which was also used in (Hu et al., 2004; Qiu et al., 2011). They have labeled opinion targets. The detailed information can be found in (Hu et al., 2004).

Domain	Language	#Sentence	#Reviews
Camera	Chinese	2075	137
Car	Chinese	4783	157
Laptop	Chinese	1034	56
Phone	Chinese	2644	123

(a) *COAE2008 dataset2*

Domain	Language	#Sentence	#Reviews
Hotel	English	1,855,351	185,829
MP3	English	289,931	30,837
Restaurant	Chinese	1,683,129	395,124

(b) *Large*

Table 2: Experimental Data Sets, # denotes the size of the reviews/sentences

In experiments, each review is segmented into sentences according to punctuations. Then sentences are tokenized and the part-of-speech of

² <http://ir-china.org.cn/coae2008.html>

³ <http://sifaka.cs.uiuc.edu/~wang296/Data/index.html>

⁴ <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>