# Postprint: Automatic Feature-Level Sentiment Word Polarity Determination Based on Modified Pointwise Mutual Information

**Authors:** Nie Hui, Shou Huanrong

**Date:** 2023-04-01T00:00:00+00:00

## Abstract

[Purpose/Significance] Corpus-based sentiment word discovery infers sentiment word polarity based on sentence context, which can significantly improve sentiment analysis accuracy and holds important application value in domain-oriented feature-level sentiment analysis tasks. [Method/Process] This paper investigates the problem of feature-level sentiment polarity judgment and proposes an automatic discrimination algorithm for the sentiment polarity of "feature-sentiment" pairs based on Pointwise Mutual Information (PMI). The algorithm leverages large-scale domain corpora, utilizing the co-occurrence relationship between opinion-expressing "feature-sentiment" pairs and seed words with clear sentiment semantics, while simultaneously introducing dependency parsing to analyze sentiment transitions between sentences. By modifying the classical PMI algorithm, it predicts the positive/negative orientation of user opinion expressions under contextual constraints. [Results/Conclusion] Experiments demonstrate that the modified algorithm significantly outperforms dictionary matching algorithms and classical PMI-based sentiment recognition algorithms. It can not only infer the sentiment orientation of opinion expressions not included in dictionaries but also accurately infer the polarity of sentiment words in "context". On two evaluation corpora—restaurant reviews and digital product reviews—the modified algorithm achieves macro-averaged F1 scores of 0.827 and 0.878, respectively. The algorithm is supported by domain-related large-scale corpora and based on probabilistic statistics and syntactic analysis. Due to convenient data acquisition, high algorithmic efficiency, good portability, and universality, it is particularly suitable for domain-oriented sentiment analysis tasks.

## Full Text

# Automatic Identification of Feature-Level Sentiment Word Polarity Based on Modified Pointwise Mutual Information

**Nie Hui, Shou Huanrong**
School of Information Management, Sun Yat-sen University, Guangzhou 510275

**Abstract:**
[**Purpose/Significance**] Corpus-based sentiment word discovery infers sentiment word polarity from contextual usage, significantly improving sentiment analysis accuracy and holding important application value for feature-level sentiment analysis tasks in specific domains. [**Method/Process**] This study investigates feature-level sentiment polarity identification, proposing an automatic polarity discrimination algorithm for "feature-opinion" pairs based on Pointwise Mutual Information (PMI). Leveraging large-scale domain corpora, the algorithm exploits co-occurrence relationships between "feature-opinion" expressions and seed words with clear sentiment semantics, while incorporating dependency parsing to capture sentiment reversals across clauses. By modifying the classical PMI algorithm, the method predicts the sentiment orientation of user opinions under contextual constraints. [**Results/Conclusion**] Experiments demonstrate that the modified algorithm significantly outperforms both dictionary-matching methods and classical PMI-based sentiment identification. It can not only infer the sentiment orientation of opinion expressions absent from existing dictionaries but also accurately determine sentiment word polarity within specific "contexts." On two evaluation corpora—restaurant reviews and digital product reviews—the macro-averaged F1 scores of the modified algorithm reach 0.827 and 0.878, respectively. Supported by domain-specific large-scale corpora and grounded in statistical analysis and syntactic parsing, the algorithm offers high efficiency, portability, and universality due to convenient data acquisition, making it particularly suitable for domain-oriented sentiment analysis tasks.

**Keywords:** sentiment analysis; pointwise mutual information algorithm; domain-specific sentiment words; context
**Classification Number:** P393.092 O212.1
**DOI:** 10.13266/j.issn.0252-3116.2020.05.012

---

The rapid development of Web technology and e-commerce is profoundly transforming people' s lives. Consumers now routinely shop online and frequently share their shopping experiences on platforms, generating vast quantities of online reviews. These user-generated product reviews, known as electronic word-of-mouth in business, have attracted considerable attention from both industry and academia due to their substantial commercial value. Extracting valuable information from online reviews constitutes a critical component of "review mining," which analyzes subjective texts containing opinions.

In practical applications, many emotionally charged terms exhibit polarity shifts depending on "context" (here referring primarily to semantic constraints from the surrounding text, such as the object being modified). Words like "big," "small," "high," and "low" exemplify this phenomenon. General-purpose dictionaries typically exclude such terms or subdivide them further. For instance, the Chinese Sentiment Lexicon Ontology from Dalian University of Technology [1] categorizes sentiment words into four types: positive, negative, dual-polarity (both positive and negative), and neutral. While positive and negative words have clear sentiment orientations, the polarity of neutral and dual-polarity words depends on "context." For feature-level sentiment analysis, accurately determining the sentiment polarity of context-dependent words is essential for ensuring precision and improving analysis quality.

The fundamental task is to determine users' sentiment orientation—positive or negative—toward specific product features. This sentiment analysis task is product-feature oriented, where sentiment word polarity often correlates with the evaluation target. For example, in smartphone reviews, the adjective "high" conveys positive sentiment in "high cost-performance ratio" but negative sentiment in "high power consumption." In stock market commentary, "high inflation rate" is negative, whereas "high interest rate" is positive. These examples demonstrate that the sentiment orientation of certain opinion words changes with their associated objects, creating what is known as "sentiment ambiguity."

This study addresses the polarity ambiguity problem in sentiment words, arguing that resolving the indeterminate sentiment orientation of dual-polarity or neutral words requires considering sentiment words and their associated objects (features) as integrated units while incorporating contextual constraints from the surrounding sentences into the analysis model. Therefore, we emphasize polarity identification for "feature-opinion" pairs, utilizing large-scale domain corpora and drawing upon PMI algorithms. By introducing dependency parsing to capture contextual constraints in sentiment expressions, we infer the polarity of "feature-opinion" pairs. The algorithm adopts an unsupervised strategy to minimize manual intervention while relying on domain corpora to ensure flexibility.

---

## 2 Related Research

### 2.1 Feature-Level Sentiment Analysis for Reviews

Feature-level sentiment analysis represents fine-grained sentiment analysis. Consider a smartphone review: "This phone' s heat dissipation is terrible—it gets hot after brief use. The camera is just average, but the overall operating speed is decent, making it recommendable for its price range." Feature-level sentiment analysis must extract user opinions like "heat dissipation—terrible," "camera—average," and "operating speed—decent," while accurately determining their sentiment orientation.

Early research by M.Q. Hu et al. [2] and B. Liu et al. [3] employed association rule mining to extract product features and used WordNet to expand adjectives for identifying user opinions (sentiment words), thereby predicting sentiment orientation toward product features. A.M. Popescu et al. [4] improved "feature-opinion" pair identification by introducing syntactic rules. Beyond effective extraction, another critical challenge is sentiment value computation, which typically relies on sentiment dictionaries. In Chinese, HowNet and the Dalian University of Technology Chinese Sentiment Lexicon Ontology [1] are widely used general-purpose sentiment dictionaries. However, these manually constructed dictionaries have limited coverage. Practical applications require expanding the lexicon to build domain-specific sentiment dictionaries. Representative work by P.D. Turney et al. [5] proposed using sentiment seed words to compute mutual information between terms for inferring the orientation of expansion words. PMI methods have been widely adopted in sentiment analysis, with sentiment dictionaries playing crucial roles.

However, researchers [6-8] have observed that sentiment words may exhibit different orientations under different interpretations. For instance, "long" carries opposite polarity when describing battery life versus meeting duration [9]. Similarly, "high" shows polarity ambiguity when evaluating "cost-performance ratio" versus "price." X. Ding et al. [6] specifically noted that for feature-level sentiment analysis, sentiment word polarity must correspond with the modified feature to ensure accuracy. Consequently, many studies have focused on accurately determining feature-dependent sentiment word polarity. Y.F. Wu et al. [8] proposed a commonsense-based disambiguation method for 14 Chinese high-frequency feature-dependent adjectives (e.g., "big, small, high, low" ), using fixed patterns with adverbs like "a bit" or "so" to determine orientation. A. Balahur et al. [10] employed majority voting across three strategies—supervised learning with context, web query hits, and rule-based methods—to determine "feature-opinion" pair polarity. H.H. Lek and D.C.C. Poo [11] clustered feature words first, then merged synonyms of sentiment and feature words using WordNet before assigning polarity through classification. Y.Q. Xia et al. [12] conducted detailed analysis of internal sentence elements, using Bayesian classification with modifying adverbs and clearly polarized sentiment words to determine polarity for feature-dependent sentiment words.

Furthermore, feature-level sentiment analysis emphasizes fine-grained sentiment computation, suggesting that quantifying user opinions should consider not only polarity but also modifying components. Expressions like "especially expensive," "not very friendly," and "fairly satisfied" demonstrate how modifiers strengthen, weaken, or even reverse sentiment polarity. Ignoring adverbial impact reduces accuracy. C. Whitelaw et al. [13] combined WordNet with semi-automatic methods to build both sentiment and modifier dictionaries, computing opinion sentiment values based on modifier attributes. S.L. Huang et al. [14] located "feature-opinion" pairs through part-of-speech patterns and determined polarity by incorporating adverbs and negation words. Shi Wei et al. [15] explored adverb functions in depth using weighted principles for refined sentiment scoring. The authors [16] previously utilized Stanford Parser to obtain modification relationships between adverbs and sentiment words, determining opinion polarity through predefined collocation patterns for fine-grained analysis.

In summary, feature-level sentiment analysis emphasizes accurate identification of user opinions, which exhibit richer forms and connotations than isolated sentiment words. It requires unified consideration of "feature-opinion" pairs alongside analysis of syntactic structure and semantic relationships in the containing sentences.

### 2.2 Sentiment Word Polarity Identification and Dictionary Construction

This study focuses on polarity identification for "feature-opinion" pairs, with the core task being sentiment word recognition and polarity determination based on modified objects—essential for dictionary construction.

**2.2.1 Dictionary-Based Sentiment Word Discovery and Polarity Identification** Dictionary expansion represents an important method for automatic sentiment dictionary generation. The fundamental approach involves: first establishing a set of sentiment seed words with clear polarity (positive or negative), then searching for synonyms/antonyms in lexical resources, expanding the seed set iteratively until convergence. M.Q. Hu and B. Liu used this method with WordNet to generate the widely adopted OpinionLexicon [2]. A. Esuli and F. Sebastiani [17] argued that terms with consistent sentiment orientation likely share similar annotation information, suggesting WordNet gloss classification could aid polarity determination. J. Kamps et al. [18] constructed a synonym network based on WordNet, inferring candidate term sentiment orientation through network distance from "good" or "bad." A. Hassan and D. Radev [19] employed advanced Markov random walk algorithms on WordNet synonym and hypernym graphs, computing average steps to polarity word sets for orientation assignment. These approaches typically treat candidate terms as independent units, inferring polarity through association with clearly semantic seeds. However, dictionary-based polarity identification is constrained by lexical resources and cannot resolve polarity shifts due to modification objects or context, performing well for semantically clear terms but potentially erring on dual-polarity or neutral words.

**2.2.2 Corpus-Based Sentiment Word Mining and Polarity Identification** Since dictionary expansion is limited by lexical resources, and many studies target specific domains, mining sentiment words from domain corpora offers greater practical value. These methods discover sentiment words from domain corpora to build specialized dictionaries and improve analysis accuracy. N. Oliveira et al. [20] annotated a stock tweet corpus, computing association between candidate terms and "bull/bear market" terms to identify domain-specific sentiment words, requiring prior corpus annotation. S. Deng et al. [21]

proposed building domain dictionaries from unlabeled data, screening sentiment candidates and using PMI to infer polarity. This approach leverages large-scale unlabeled domain corpora for efficient modeling and improved dictionary effectiveness. Xi Yahui [22] constructed domain sentiment dictionaries in two stages: first building word matrices using PMI and contextual constraints, then applying label propagation to expand seed dictionaries, with domain association determined by sentiment conflict frequency. K. Labille et al. [23] built sentiment dictionaries for 15 domains using probability theory and information theory, demonstrating excellent performance. Deep learning methods have also been applied: Y. Li et al. [24] incorporated prior knowledge into word embedding training for domain adaptation, while Lin Jianghao et al. [25] used large-scale corpus-trained Word2Vec representations combined with sentiment dictionaries for deeper word representation.

Overall, corpus-based approaches suit domain-dependent sentiment word identification, with unlabeled corpora offering practical feasibility. However, domain dictionary construction alone cannot resolve context-dependency for dual-polarity or neutral words. This study defines "context" specifically as contextual constraints in sentiment word sentences, particularly the modified features. We argue that accurate polarity inference for such words requires incorporating both contextual constraints and modified objects into the model. Unlike dictionary-building approaches, this study treats "feature-opinion" pairs as the unit of analysis, performing polarity inference on complete opinion expressions while modifying the PMI algorithm to incorporate "sentiment reversal" states obtained through dependency parsing, thereby addressing sentiment ambiguity and improving accuracy.

---

# 3 Polarity Identification for "Feature-Opinion" Pairs Based on PMI

## 3.1 Mutual Information and PMI-Based Sentiment Polarity Identification

Mutual information, a concept from information theory, measures word correlation in text mining. PMI calculation is based on co-occurrence probability—higher co-occurrence rates indicate stronger correlation. For words $w_1$ and $w_2$:

$$PMI(w_1, w_2) = log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

where $P(w_1)$ and $P(w_2)$ are individual occurrence probabilities, and $P(w_1, w_2)$ is co-occurrence probability. P.D. Turney and M.L. Littman [5] introduced mutual information to sentiment analysis as Semantic Orientation PMI (SO-PMI), based on the observation that words with similar sentiment orientation frequently co-occur (e.g., "happy-joyful," "pain-sorrow"), while opposite-orientation words do not. Thus, for a candidate sentiment word w, comparing its PMI with positive seed set Pos_{set} versus negative seed set Neg_{set} can infer its polarity:

$$SO_{PMI}(w) = \sum_{w_P \in Pos\_set} PMI(w, w_P) - \sum_{w_N \in Neg\_set} PMI(w, w_N)$$

If SO-PMI(w) > 0, w tends toward positive polarity; if SO-PMI(w) < 0, negative. While classical PMI-based sentiment computation is theoretically sound, real language environments are more complex. Words with opposite polarity can co-occur, as in "pain and happiness" or "joy begets sorrow," causing simple SO-PMI calculation to deviate.

Moreover, this study focuses on "feature-opinion" pairs rather than isolated sentiment words. Sentiment consistency based on co-occurrence varies with syntactic structure. Consider: "This restaurant' s price is reasonable, but the service is terrible." The review evaluates two aspects: "price-reasonable" (positive) and "service-terrible" (negative). Despite co-occurrence, the polarity of the target "feature-opinion" reverses due to the conjunction "but." Similar sentiment co-occurrence with reversal is common, where transitional relationships create semantic differences despite shared subjects, as in: "Price reasonable, service terrible," "Price reasonable, service not terrible," "Price unreasonable, service terrible," and "Price unreasonable, but service not terrible." These four sentences exhibit clear semantic and sentiment differences. In pure PMI calculation considering only co-occurrence, all four contain "price-reasonable" and "service-terrible," yielding equivalent PMI values but obviously incorrect judgments. We argue that polarity inference for "feature-opinion" pairs should incorporate both co-occurrence and sentiment reversal relationships, using dependency parsing to extract more accurate sentiment relationships and reduce misclassification.

### 3.2 Modified PMI Algorithm for "Feature-Opinion" Polarity Identification

This study improves upon Turney' s [5] SO-PMI method by employing dependency parsing to introduce reversal/transition relationships, modifying sentiment consistency calculation between "feature-opinion" pairs and seed words to reduce identification errors. Through observation, we analyzed sentiment reversal phenomena in corpora, with examples shown in Table 1 .

**Table 1** Impact of Negation/Transition Words on Sentence Sentiment Semantics (Examples)

| Sentence | Sentiment Polarity Consistency | Reversal Relationship |
|---|---|---|
| Price reasonable, but service terrible | Inconsistent | Service-terrible → Reverse Price-reasonable: Positive |
| Price unreasonable, but service not terrible | Inconsistent | Service-terrible → Reverse Price-unreasonable: Negative |
| Price unreasonable, service also terrible | Consistent | Service-terrible → Reverse Price-unreasonable: Negative |
| Price reasonable, service not terrible | Consistent | Service-terrible → Reverse Price-reasonable: Positive |

Note: Assuming "terrible" as negative seed, "price-reasonable" and "price-unreasonable" as target feature-opinion pairs.

In Table 1, both "but" and "not" indicate sentiment reversal. Within the co-occurrence window, when reversal words appear an odd number of times, the "feature-opinion" polarity becomes inconsistent with the seed word polarity, triggering sentiment reversal. Therefore, the polarity of "price-reasonable" inferred from its co-occurrence with "service-terrible" must be adjusted according to reversal status.

**Algorithm Description:**

Let Corpus be the analysis corpus, from which "feature-opinion" pairs <f, o> are extracted using feature-opinion extraction algorithms. Let PMI_{pos}(<f, o>) and PMI_{neg}(<f, o>) represent the association degrees between <f, o> and positive/negative sentiment, respectively. Let Pos_{set} and Neg_{set} be positive and negative seed word sets. Let w be a sentiment seed word, w Neg_{set} Pos_{set}.

**Input:** <f, o>, Corpus, Pos_{set}, Neg_{set}
**Output:** Sentiment polarity of <f, o>

**Step 1:** Initialize PMI_{pos}(<f, o>) = 0, PMI_{neg}(<f, o>) = 0 for all <f, o> Corpus.

**Step 2:** Segment original sentences in the corpus, setting the co-occurrence window as segmented short clauses containing <f, o> and a seed word w.

**Step 3:** Introduce a negation/transition dictionary and apply dependency parsing to short clauses to detect sentiment reversal for w.

**Step 4:** For w  Pos_{set}, compute PMI(<f, o>, w). If no sentiment reversal occurs in the clause:

$$PMI_{pos}(<f,o>) = PMI_{pos}(<f,o>) + PMI(<f,o>,w)$$

If reversal occurs:

$$PMI_{neg}(<f,o>) = PMI_{neg}(<f,o>) + PMI(<f,o>,w)$$

**Step 5:** For w  Neg_{set}, compute PMI(<f, o>, w). If no sentiment reversal occurs:

$$PMI_{neg}(<f,o>) = PMI_{neg}(<f,o>) + PMI(<f,o>,w)$$

If reversal occurs:

$$PMI_{pos}(<f,o>) = PMI_{pos}(<f,o>) + PMI(<f,o>,w)$$

**Step 6:** Compute the SO-PMI for <f, o>:

$$SO_{PMI}(<f,o>) = PMI_{pos}(<f,o>) - PMI_{neg}(<f,o>)$$

If SO_{PMI}(<f, o>) > 0, <f, o> is more associated with positive sentiment, likely expressing positive opinion; otherwise, negative.

Notably, Step 2 (co-occurrence window extraction) uses the SentenceSplit interface from Harbin Institute of Technology' s Language Technology Platform, which segments long sentences by punctuation marks (semicolons, exclamation points, periods). Short clauses may still contain sub-clauses separated by commas. We use short clauses as window units for dependency parsing. In Step 3, reversal detection uses dependency relations. For example, in "Price reasonable, but service terrible," the dependency analysis reveals a direct dependency between "but" and "terrible," so the reversal triggered by "but" is attributed to the negative word "terrible." Thus, within the window, "price-reasonable" has its polarity adjusted relative to "terrible" based on reversal status.

This algorithm directly infers sentiment orientation of "feature-opinion" pairs from corpora. Treating "feature-opinion" as a unified entity affected by contextual constraints within co-occurrence windows, dependency parsing enhances judgment accuracy.

---

# 4 Experiments and Analysis

## 4.1 Experimental Design and Process

The experimental workflow is illustrated in Figure 1 [Figure 1: see original paper], comprising four components: data collection and preprocessing, seed sentiment dictionary and negation/transition dictionary construction (resource building), automatic sentiment polarity identification for “feature-opinion” pairs based on improved PMI, and algorithm performance evaluation.

**Figure 1** Experimental Design and Process

### 4.2 Experimental Corpus and Resource Construction

**4.2.1 Experimental Corpus** We selected digital product and restaurant domains, collecting real review data. Restaurant reviews were crawled from Dianping.com, targeting Cantonese restaurants in Guangzhou based on rankings. The crawler collected user review content and star ratings (spanning July 7, 2004 to October 31, 2017), yielding 400,126 reviews: 23,663 negative (rating $<$ 3) and 376,463 positive (rating $>$ 3). Digital product reviews were obtained from JD.com via a publicly shared dataset (https://download.csdn.net/download/qq_{30365565}/10657579), containing 108,919 negative and 101,905 positive reviews after cleaning and deduplication.

**4.2.2 Dictionaries** We integrated four external dictionaries to create a general sentiment lexicon (https://download.csdn.net/download/weixin_{42018090}/10358043): HowNet positive/negative sentiment words, NTUSD (National Taiwan University Sentiment Dictionary), Tsinghua University Li Jun’ s Chinese sentiment dictionary, and Dalian University of Technology Sentiment Lexicon Ontology. After merging, we removed redundant and conflicting entries. Our seed sentiment set comprised 10,653 positive and 15,888 negative words with clear sentiment semantics.

For sentiment reversal detection, we constructed a negation/transition dictionary. We compiled 38 common Chinese negation words (e.g., “not,”“no,”“cannot” ) from online resources. Sentiment transitions typically occur between adjacent clauses—prior clauses state facts, subsequent clauses express contrary meanings, often revealing the speaker’ s true intent. Chinese transition words are conjunctions with limited variety. We manually compiled 15 common transition words (Table 2 ).

**Table 2** Negation/Transition Dictionary

| Negation Words | Transition Words |
|---|---|
| 不, 没, 否定, 不能, 反而, 并非, 相反, 无法, 不够, 不是, 不曾, 未必, 不要, 未曾, 全无, 不太, 未免, 不足, 未经, 并不, 不意, 不意味, 不意味着, 不足以, 不怎么, 反过来, 并不一定, 没有, 不会, 毫无, 毫不, 并不比, 绝不, 不怎么, 难以, 不算 | 可是, 但是, 虽说, 尽管, 固然, 不过, 只是, 虽然, 却, 然而, 就是, 除了 |

**4.2.3 Evaluation Corpus** To evaluate algorithm performance, we constructed an evaluation corpus by randomly sampling 1,000 JD electronics reviews and 1,600 Dianping restaurant reviews. A data company annotated these via outsourcing, extracting "feature-opinion" pairs and labeling their sentiment orientation (positive/negative) as triples <feature, opinion, polarity>. After annotation, 10% were randomly sampled for quality control, achieving 94% accuracy. To ensure balance, positive and negative reviews were equally represented. Table 3 shows annotation examples.

**Table 3** Evaluation Corpus Annotation Examples

| Review_{ID} | Review Content | Annotation |
|---|---|---|
| … | Used for a day before reviewing. Operating speed is quite fast, but battery not durable, camera effect so-so | <operating speed, fast, 1> <battery, not durable, -1> |
| … | Phone looks great, resolution poor, lags when gaming, weak WiFi signal | <phone, great, 1> <resolution, poor, -1> <gaming, lag, -1> <signal, weak, -1> |

Note: 1 = positive sentiment, -1 = negative sentiment.

### 4.3 Experiments and Results Analysis

**4.3.1 Evaluation Metrics** We used precision (P), recall (R), and F1-score. For comprehensive evaluation, we computed macro- and micro-averages across positive/negative classes. Table 4 shows the confusion matrix, with formulas (3)-(5) defining the metrics.

**Table 4** Positive/Negative Sentiment Discrimination Confusion Matrix

| | Human Positive/Negative | Human Non-Positive/Negative |
|---|---|---|
| **Algorithm Positive/Negative** | TPp/TPn | FNp/FNn |
| **Algorithm Non-Positive/Negative** | FPp/FPn | TNp/TNn |

Positive/negative sentiment identification metrics:

$$P_p = \frac{TP_p}{TP_p + FP_p}, \quad R_p = \frac{TP_p}{TP_p + FN_n}, \quad F1_p = \frac{2 \cdot P_p \cdot R_p}{P_p + R_p}$$

$$P_n = \frac{TP_n}{TP_n + FP_n}, \quad R_n = \frac{TP_n}{TP_n + FN_n}, \quad F1_n = \frac{2 \cdot P_n \cdot R_n}{P_n + R_n}$$

Macro-averages are means of positive/negative metrics. Micro-averages are computed as:

$$P_{micro} = \frac{TP_p + TP_n}{TP_p + TP_n + FP_p + FP_n}$$

$$R_{micro} = \frac{TP_p + TP_n}{TP_p + TP_n + FN_p + FN_n}$$

$$F1_{micro} = \frac{2 \cdot P_{micro} \cdot R_{micro}}{P_{micro} + R_{micro}}$$

**4.3.2 Experiments and Results** Implemented in Python, we conducted two experiments.

**Experiment 1: Impact of Corpus Balance on Performance.** Reviews were divided by star ratings into positive and negative sets. Preliminary experiments revealed that distribution balance significantly affected performance. We introduced balance parameter K (Formula 6) to examine its effect.

$$K = \frac{|Pos\_review|}{|Neg\_review|}$$

where Pos_{review} and Neg_{review} are positive/negative review subsets. Table 5 shows corpus quantities, Table 6 and Table 7 present results.

**Table 5** Positive/Negative Corpus Quantities (in reviews)

| Corpus | Positive | Negative |
|---|---|---|
| Dianping Restaurant | 376,463 | 23,663 |
| JD Electronics | 101,905 | 108,919 |

**Table 6** Corpus Balance Experiment Results (Dianping Restaurant Reviews)

| K | 9:1 | 5:1 | 1:1 | 1:5 | 1:9 |
|---|---|---|---|---|---|
| P_{macro} | 0.793 | 0.866 | 0.827 | 0.708 | 0.699 |
| R_{macro} | 0.753 | 0.786 | 0.791 | 0.682 | 0.677 |
| F1_{macro} | 0.773 | 0.826 | 0.827 | 0.695 | 0.687 |
| P_{micro} | 0.794 | 0.862 | 0.862 | 0.702 | 0.694 |
| R_{micro} | 0.765 | 0.779 | 0.803 | 0.699 | 0.694 |
| F1_{micro} | 0.779 | 0.820 | 0.832 | 0.700 | 0.698 |

**Table 7** Corpus Balance Experiment Results (JD Electronics Reviews)

| K | 9:1 | 5:1 | 1:1 | 1:5 | 1:9 |
|---|---|---|---|---|---|
| P_{macro} | 0.841 | 0.874 | 0.878 | 0.819 | 0.784 |
| R_{macro} | 0.811 | 0.824 | 0.856 | 0.794 | 0.771 |
| F1_{macro} | 0.825 | 0.849 | 0.878 | 0.806 | 0.777 |
| P_{micro} | 0.871 | 0.895 | 0.907 | 0.854 | 0.829 |
| R_{micro} | 0.862 | 0.879 | 0.898 | 0.849 | 0.825 |
| F1_{micro} | 0.867 | 0.887 | 0.907 | 0.852 | 0.827 |

Results show optimal performance at balanced distribution (K = 1:1): F1_{macro} = 0.878, F1_{micro} = 0.907 on JD corpus; F1_{macro} = 0.827, F1_{micro} = 0.832 on Dianping corpus. Performance degrades with skewed distributions. PMI relies on co-occurrence—words with similar sentiment orientation frequently co-occur. If "feature-opinion" pairs fail to establish relationships with sentiment-certain words within windows, polarity cannot be determined. Insufficient co-occurrence support leads to judgment errors. With fewer negative reviews (K = 9:1), negative "feature-opinion" pairs lack adequate co-occurrence relationships, impairing precision and recall (P_n = 0.691, R_n = 0.652 on JD; P_n = 0.689, R_n = 0.551 on Dianping). Based on these results, subsequent experiments use K = 1:1.

**Experiment 2: Sentiment Polarity Identification for "Feature-Opinion" Pairs Using PMI.** We ran the improved SO-PMI on balanced corpora from both domains, evaluating against the annotated test set and comparing with dictionary matching and standard SO-PMI. Results are shown in Figure 2 [Figure 2: see original paper] and Figure 3 [Figure 3: see original paper].

**Figure 2** Algorithm Performance Comparison (JD Digital Product Reviews)

On JD reviews, the improved algorithm achieved F1_{macro} = 0.878 and F1_{micro} = 0.907, with highest recall among the three methods. Dictionary matching achieved highest precision (P_{macro} = 0.946, P_{micro} = 0.975) but lowest recall (R_{macro} = 0.307, R_{micro} = 0.350), as it cannot judge unlisted sentiment words. Standard PMI showed slightly lower precision (P_{macro} = 0.869, P_{micro} = 0.839) but significantly higher recall (R_{macro} = 0.710, R_{micro} = 0.835), demonstrating stronger discovery capability. The improved PMI further enhanced both precision and recall, achieving the best overall performance.

**Figure 3** Algorithm Performance Comparison (Dianping Restaurant Reviews)

Similar results were observed on restaurant reviews, with the improved algorithm performing best.

Performance improvements stem from three factors:

1. **Large-scale balanced corpora overcome dictionary limitations.** The improved algorithm targets "feature-opinion" pairs where "opinion"

extends beyond dictionary-listed sentiment words to broader emotional expressions like "awesome," "after-sales service is good," "getting cheaper," and "not great." These unlisted expressions can be inferred, significantly boosting recall.

2. **Contextual reversal constraints improve precision.** The improved algorithm achieves higher precision than standard SO-PMI because the latter relies solely on word co-occurrence without considering feature-opinion dependencies or contextual reversals. By using dependency parsing to analyze reversal relationships, the algorithm more accurately captures sentiment co-occurrence consistency. For example, in "The soup is too salty, but the vegetables are delicious," standard SO-PMI would incorrectly infer "taste-salty" as positive due to co-occurrence with "delicious," while the improved algorithm correctly identifies it as negative.

3. **Context-aware identification of dual-polarity words enhances overall effectiveness.** Neutral and dual-polarity words like "big," "small," "high," "low" require modification objects to convey accurate sentiment. Our algorithm analyzes "feature-opinion" pairs, considering opinion dependency on features. With sufficient corpora, it accurately identifies sentiment orientation under contextual constraints. Table 8 shows sample results for "big" on JD corpus.

**Table 8** Sample "Feature-Opinion" Pairs with Adjective "Big"

| Review_{ID} | Feature-Opinion Pair | Polarity |
|---|---|---|
| ··· | screen-big | Positive |
| ··· | memory space-big | Positive |
| ··· | fan sound-big | Negative |
| ··· | panel gap-big | Negative |
| ··· | photo noise-big | Positive (Error) |

Empirically, "big" is positive for "memory space" but negative for "fan sound." The algorithm makes appropriate judgments based on modified objects. However, errors occur: "photo noise-big" was misjudged as positive because "photo noise" appeared only once in a sentence with "clear" and "average" as sentiment guides not in the seed dictionary, making "first-class" (positive) the closest sentiment word. This demonstrates that corpus balance and seed dictionary completeness significantly impact performance. Nevertheless, analyzing nearly 200,000 balanced reviews achieved F1 near 0.9, substantially outperforming dictionary matching and standard PMI.

## 5 Conclusion and Future Research

This study optimizes PMI-based sentiment computation using large-scale domain corpora for automatic feature-level sentiment polarity identification. The algorithm infers polarity of target "feature-opinion" pairs based on co-occurrence with sentiment-clear seed words while considering contextual reversal states from dependency parsing. Experiments on real corpora demonstrate that the proposed SO-PMI optimization outperforms general dictionary matching and classical SO-PMI, operating without dictionary constraints and applicable to various domain-specific sentiment analysis tasks.

Contributions are threefold: (1) Analyzing "feature-opinion" pairs partially resolves sentiment ambiguity—e.g., inferring "big' s" orientation based on modified objects, enabling context-specific sentiment analysis with sufficient corpora. (2) The algorithm enhances classical SO-PMI by extracting intra-sentence co-occurrence relationships and analyzing reversals via dependency parsing, accurately capturing sentiment consistency and improving precision. (3) As a corpus-based method, it requires minimal manual intervention, no training process, and offers convenient data acquisition, strong universality, and easy migration across domains.

Notably, experiments across two domains achieved favorable results, with slightly better performance on experience products (restaurants) than search products (electronics). Future research will expand to other domains (stock reviews, book reviews, etc.) for deeper investigation.

## References

[1] Dalian University of Technology Information Retrieval Laboratory. Dalian University of Technology Chinese Sentiment Lexicon Ontology [EB/OL]. [2019-01-10]. http://ir.dlut.edu.cn/EmotionOntologyDownload.

[2] HU M Q, LIU B. Mining and summarizing customer reviews [C] // Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM Press, 2004: 168-177.

[3] LIU B, HU M Q, CHENG J S. Opinion observer: analyzing and comparing opinions on the Web [C] // Proceedings of the 14th international conference on WWW. New York: ACM Press, 2005: 342-351.

[4] POPESCU A M, ETZIONI O. Extracting product features and opinions from reviews [C] // Natural language processing and text mining. London: Springer, 2007: 9-28.

[5] TURNEY P D, LITTMAN M L. Measuring praise and criticism: inference of semantic orientation from association [J]. ACM transactions on information systems, 2003, 21(4): 315-346.

ChinaRxiv 中X

[6] DING X, LIU B, YU P S. A holistic lexicon-based approach to opinion mining [C] // Proceedings of the 2008 international conference on Web search and data mining. New York: ACM Press, 2008: 231-240.

[7] DRACUT E C, YU C, SISTLA P A, et al. Construction of a sentimental word dictionary [C] // Proceedings of the 19th ACM international conference on information and knowledge management. New York: ACM Press, 2010: 1761-1764.

[8] WU Y F, WEN M M. Disambiguating dynamic sentiment ambiguous adjectives [C] // Proceedings of the 23rd international conference on computational linguistics. Stroudsburg: Association for Computational Linguistics, 2010: 1191-1199.

[9] WANG K, XIA R. A survey on automatic construction methods of sentiment lexicons [J]. Acta automatica sinica, 2016, 42(4): 495-509.

[10] BALAHUR A, MONTOYO A O. Applying opinion mining techniques for the disambiguation of sentiment ambiguous adjectives in SemEval-2 task 18 [C] // Proceedings of the 5th international workshop on semantic evaluation. Stroudsburg: Association for Computational Linguistics, 2010: 444-447.

[11] LEK H H, POO D C C. Sentix: an aspect and domain sensitive sentiment lexicon [C] // Proceedings of the 2012 IEEE 24th international conference on tools with artificial intelligence. Washington, DC: IEEE Computer Society, 2012: 261-268.

[12] XIA Y Q, CAMBRIA E, HUSSAIN A, et al. Word polarity disambiguation using Bayesian model and opinion-level features [J]. Cognitive computation, 2014, 7(3): 369-380.

[13] WHITELAW C, GARG N, ARGAMON S. Using appraisal groups for sentiment analysis [EB/OL]. [2019-10-01]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.8147&rep=

[14] HUANG S L, CHENG W C. Discovering Chinese sentence patterns for feature-based opinion summarization [J]. Electronic commerce research and application, 2015, 14(6): 582-591.

[15] SHI W, WANG H W, HE S Y. Semantic-based sentiment analysis of Chinese online reviews [J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(8): 860-867.

[16] NIE H. Aggregation and quantification of product review opinions under latent topic models [J]. Journal of information science, 2017, 36(6): 565-573.

[17] ESULI A, SEBASTIANI F. Determining the semantic orientation of terms through gloss classification [C] // Proceedings of the 14th ACM international conference on information and knowledge management. New York: ACM Press, 2005: 617-624.

[18] KAMPS J, MARX M, MOKKEN R J, et al. Using WordNet to measure semantic orientations of adjectives [EB/OL]. [2019-04-11].

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.2534&rep=rep1&type=pdf.

[19] HASSAN A, RADEV D. Identifying text polarity using random walks [C] // Proceedings of the 48th annual meeting of the association for computational linguistics. Stroudsburg: Association for Computational Linguistics, 2010: 395-403.

[20] OLIVEIRA N, CORTEZ P, AREAL N. Stock market sentiment lexicon acquisition using microblogging data and statistical measures [J]. Decision support systems, 2016, 85: 62-73.

[21] DENG S, SINHA A P, ZHAO H. Adapting sentiment lexicons to domain-specific social media texts [J]. Decision support systems, 2017, 94: 65-76.

[22] XI Y H. Construction of domain sentiment lexicon for product reviews [J]. Journal of Chinese information processing, 2016, 30(5): 136-144.

[23] LABILLE K, GAUCH S, ALFARHOOD S. Creating domain-specific sentiment lexicons via text mining [EB/OL]. [2019-04-11]. http://www.sentric.net/wisdom2017labille.pdf.

[24] LI Y, PAN Q, YANG T, et al. Learning word representations for sentiment analysis [J]. Cognitive computation, 2017, 9(6): 843-851.

[25] LIN J H, ZHOU Y M, YANG A M, et al. Domain-specific sentiment lexicon construction based on word vectors [J]. Journal of Shandong University (engineering science), 2018, 48(3): 40-47.

---

**Author Contributions:**
Nie Hui: Research problem formulation, framework design, paper revision and refinement.
Shou Huanrong: Data collection and experimentation, initial draft writing.

*Note: Figure translations are in progress. See original paper for figures.*