Opinion Mining and Relationship Discovery Using CopeOpi Opinion Analysis System

Lun-Wei Ku, Hsiu-Wei Ho, and Hsin-Hsi Chen*
Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 106. E-mail: {lwku, xwhe}@nlg.csie.ntu.edu.tw, hhchen@ntu.edu.tw

We present CopeOpi, an opinion-analysis system, which extracts from the Web opinions about specific targets, summarizes the polarity and strength of these opinions, and tracks opinion variations over time. Objects that yield similar opinion tendencies over a certain time period may be correlated due to the latent causal events. CopeOpi discovers relationships among objects based on their opinion-tracking plots and collocations. Event bursts are detected from the tracking plots, and the strength of opinion relationships is determined by the coverage of these plots. To evaluate opinion mining, we use the NTCIR corpus annotated with opinion information at sentence and document levels. CopeOpi achieves sentence- and document-level *f*-measures of 62% and 74%. For relationship discovery, we collected 1.3M economics-related documents from 93 Web sources over 22 months, and analyzed collocation-based, opinion-based, and hybrid models. We consider as correlated company pairs that demonstrate similar stock-price variations, and selected these as the gold standard for evaluation. Results show that opinion-based and collocation-based models complement each other, and that integrated models perform the best. The top 25, 50, and 100 pairs discovered achieve precision rates of 1, 0.92, and 0.79, respectively.

Introduction

Sentiment analysis has attracted much attention in recent years because a large amount of subjective information is disseminated through various platforms on the Web such as review sites, forums, discussion groups, blogs, and news. Web users are willing to share their thoughts or feelings with others after reading books, watching movies, buying products, and so on. Consulting specific information sources and summarizing the newly discovered opinions aids governments in improving services, companies in marketing products, and customers in purchasing items.

In the past, extracted opinions have only focused on individual targets, with effects on multiple targets being left for users to interpret. Few works touch on automatic comparison of opinions about multiple targets. Opinions usually accompany specific events; this phenomenon makes event-burst detection indispensable. Given the detected events, opinion mining can further identify the relationship between opinion polarity and the correlated events: Two entities may be related if different events always result in similar opinions.

We employ the results of opinion mining in relationship discovery, and compare the results with those of the traditional collocation model, which discovers relationships among terms based on their co-occurrences in physical contexts such as documents, sentences, and adjacent words. The basic idea is that if entities involved in the same sequence of events yield similar opinion trends, they may be correlated. To minimize chance co-occurrences, two entities should be observed over a sufficient amount of time.

We hence propose the opinion-analysis system CopeOpi. This system extracts opinions, providing the information necessary for relationship discovery. To demonstrate its feasibility, listed companies are considered as the experimental targets for relationship discovery. We propose several models and use the best model in CopeOpi. For opinion-based models, we use original curves, digitized curves, and smoothened curves of tracking plots to examine the opinions' effects on relationship discovery. For collocation-based models, we extract collocations at the word, sentence, and document level, and discuss their impact on collocation degree. In total, eight models, including two collocation-based, four opinion-based, and two integration models, are evaluated. This system also includes a user interface and the necessary functions for an opinion-analysis system.

This paper is composed of six sections. In the next section, we compare this research with related works and summarize its contributions. Then we describe the applications of the CopeOpi system to Web mining and relationship discovery, and in the following section we propose several ways to implement the kernel of the system. In the Experimental Setup and Evaluation sections, we describe the experimental setup including the data sets and gold standard, and

Received June 24, 2008; revised February 9, 2009; accepted February 10, 2009

*Corresponding author

© 2009 ASIS&T • Published online 7 April 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21067

then discuss the experimental results for opinion mining and relationship discovery, respectively. Finally, we conclude.

Related Work

Opinion extraction, which identifies subjective information from designated sources, is fundamental for opinion summarization and tracking (Ku, Liang, & Chen, 2006). In the past, researchers have proposed opinion extraction for different granularities such as words, sentences, and documents. At the word level, Wiebe (2000) learned subjective adjectives from corpora, Riloff, Wiebe, and Wilson (2003) learned subjective nouns, and Takamura, Inui, and Okumura (2005) extracted opinion polarities for words. Wilson, Wiebe, and Hoffmann (2005) recognized contextual polarities of phrase-level units. Beyond the word level, Riloff and Wiebe (2003) learned extraction patterns for subjective expressions; Kim and Hovy (2004) further determined the polarities of the expressions; and Jindal and Liu (2006) even identified comparative sentences to measure the strength of opinions. At the document level, various applications and techniques have been proposed for opinion extraction (Dave, Lawrence, & Pennock, 2003; Pang, Lee, & Vaithyanathan, 2002; Wiebe, Wilson, & Bell, 2001).

The three important roles in an opinion are the opinion holder, the opinion target, and the opinion itself. For the opinion itself, we should additionally determine its polarity. The simplest operation of opinion extraction is binary decision, such as opinion versus nonopinion, or positive versus negative. Pang et al. (2002) and Alm, Roth, and Sproat (2005) use machine-learning techniques for classification, while Takamura et al. (2005) and Ghose, Ipeirotis, and Sundararajan (2007) adopt regression models. Riloff and Wiebe (2003), and Choi, Cardie, Riloff, and Patwardhan (2005) employ linguistic or structural information as auxiliary cues. In addition to classification, opinion holders can be extracted (Breck, Choi, & Cardie, 2007; Choi et al.; Kim & Hovy, 2005), and their opinions can be summarized (Hu & Liu, 2004a; Ku, Lee, Wu, & Chen, 2005; Seki, Eguchi, & Kando, 2005; Stoyanov & Cardie, 2006) and even tracked (Ku et al., 2006). Opinions can be mined from texts of different genres, such as product reviews (Bai, Padman, & Airoldi, 2005; Dave et al., 2003; Hu & Liu, 2004a, 2004b; Liu, Huang, An, & Yu, 2007; Morinaga, Yamanishi, Tateishi, & Fukushima, 2002; Pang & Lee, 2005), news documents (Ku et al., 2006), blog articles (Ku et al., 2006; Mei, Ling, Wondra, Sum, & Zhai, 2007; Yang, Yu, Valerio, Zhang, & Ke, 2007), and so on. Many interesting applications (Liu, Hu, & Cheng, 2005; Hu & Liu, 2004a; Ku et al., 2007a, 2007b) have been developed.

Among the opinion-analysis systems that have been developed for alphabetic languages, OASYS (Cesarano, Picariello, Reforgiato, & Subrahmanian, 2007) is the most famous. Both OASYS and CopeOpi allow users input their queries and select preferred data sources, and then track opinions in a time zone. However, OASYS analyzes opinions from each data source separately, while CopeOpi collects articles from all data sources and generates a summary. OASYS lists related documents only by publication time, but CopeOpi ranks documents according to their significance by finding important events (Ku et al., 2005). For both systems, extracting opinions is the main focus, while holders and targets are identified implicitly when retrieving relevant documents.

Relationship discovery is a hot research topic in the social-network domain. Lin and Chen (2008) mine relationships from features of entities and collocated terms from documents. Mori, Ishizuka, and Matsuo (2007) further mine predefined relationships. We mine relationships from an opinion view, that is, from comments and evaluations people offer about these entities; this is very different from other works. We have selected companies as example entities for mining relationships in CopeOpi. The relationships mined by CopeOpi from companies are financial, but are not specific template elements like *employer*, *employee*, or *ownership*, as defined in MUC (Chinchor, 1998). The company relationships are targeted more toward overall evaluations, such as stock prices. We feel these relationships are useful, as CopeOpi manages to align companies whose financial performances are correlated. Because CopeOpi does not apply any domain-specific methods to relationship discovery, it is able to mine relationships among any targets given relevant comments.

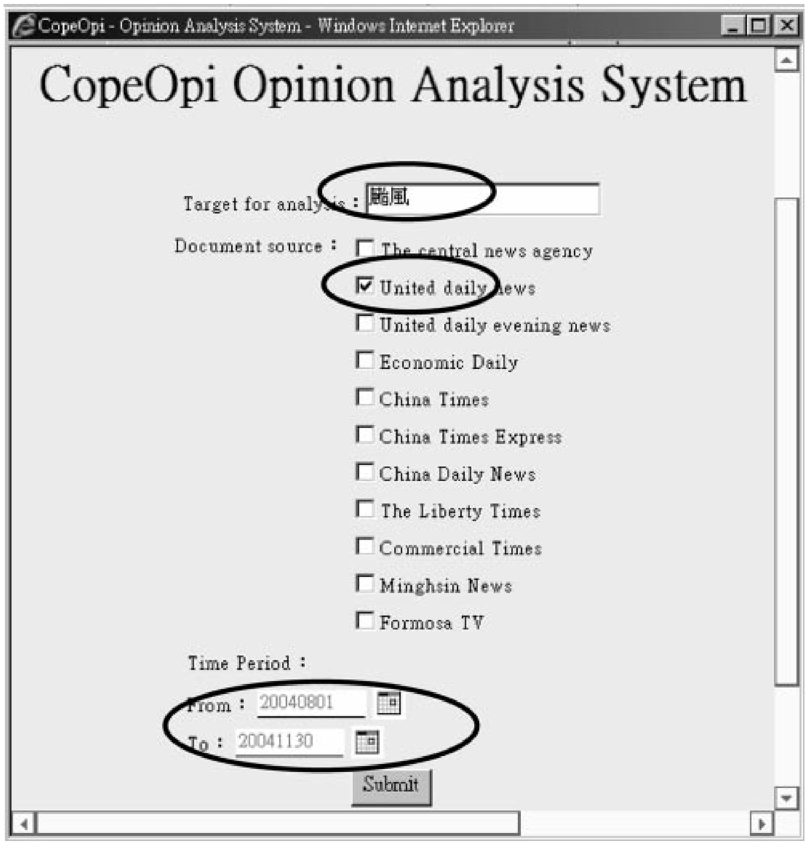
It is also important to have an effective user interface for an opinion-analysis system. CopeOpi combines graphical and text media to present complex information like opinions. It displays clear opinion polarities first using a tricolor opinion-tracking line along the timeline, and then a tricolor visualization of texts, in which the text colors indicate the individual opinion polarity of each sentence or document. OASYS also uses multicolor tracking lines, but to distinguish different data sources. Carenini's team proposed a graphical user interface for evaluative texts (Carenini, Ng, & Pauls, 2006), in which color blocks are used to present the evaluations for components of products. However, their blocks contain no time information for tracking and are more suitable for processing only reviews of specific products due to predefined features.

CopeOpi Opinion Analysis System

CopeOpi (Chinese opinion extraction system for Opinion information) supports opinion mining and relationship discovery. This section begins with an illustration of its uses. Related technologies are presented in the next section.

Opinion Mining

For opinion mining, the input is a target to be analyzed, the data sources, and a time period. Users type in the entities they are interested in, select one or more data sources, and specify a time period. Figure 1 shows an example, where *typhoon* (颱風) is specified as the target, United Daily News as the document source, and August 1, 2004, to November 30, 2004, as the time period.



CopeOpi - Opinion Analysis System - Windows Internet Explorer

CopeOpi Opinion Analysis System

Target for analysis : 颱風

Document source : □ The central news agency
☑ United daily news
□ United daily evening news
□ Economic Daily
□ China Times
□ China Times Express
□ China Daily News
□ The Liberty Times
□ Commercial Times
□ Minghsin News
□ Formosa TV

Time Period :
From : 20040801
To : 20041130

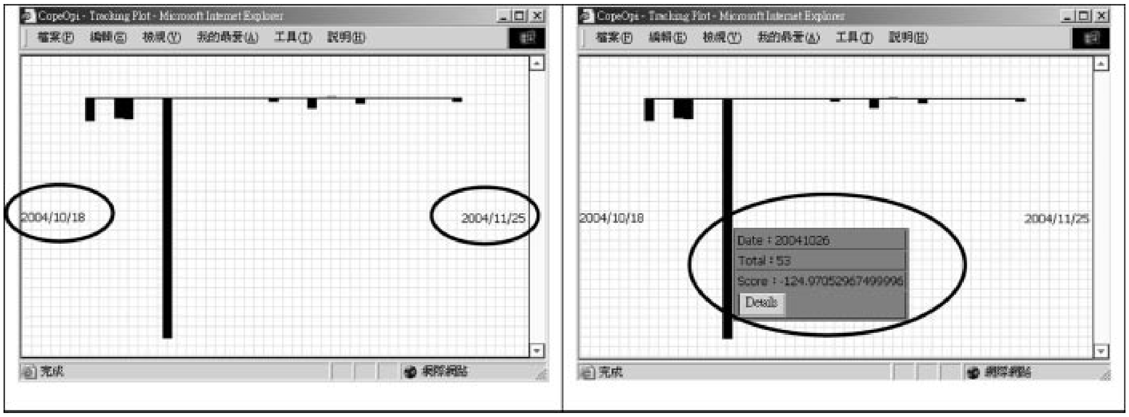
FIG. 1. CopeOpi opinion mining.

CopeOpi provides a graph-based summary to present the results of opinion mining. A bar graph summarizing the opinion score of each day in the timeline summarizes all of the opinions and their changes. We hereafter refer to this bar graph as a "tracking plot." To generate the tracking plot, each day's documents related to the target are first extracted by an IR system. Then the length of each bar in the graph is set to the sum of the opinion scores of documents for that day. In CopeOpi, the colors green, gray, and red denote positive, neutral, and negative opinions. For a clearer representation, we here replace green with bold gray and red with bold black, and there is no neutral date in this example. Figure 2 below shows the tracking plot for the *typhoon* query submitted in Figure 1.

The first and the last dates of the assigned time period are shown in the left and the right of the tracking plot. The downward bar indicates a date with strong negative opinions, while the upward bar denotes a positive one. The length of each bar indicates the combined strength of opinion for that day, with longer bars signifying stronger opinions. A positive date is a date in which the majority of opinions is positive, and vice versa. When users move the cursor to a given bar, the summary for that day is displayed, as shown in the right window of Figure 2. The date with the longest negative bar is October 26th, 2004, where a total of 53 documents were found to contain the query term *typhoon* on that day and the total opinion score is −124.97. Moreover, the number of bursts denotes the appearances of events or monitored foci. Figure 2 reflects the three major typhoons that hit Taiwan in October 2004, including No. 0418 typhoon Aere, No. 0423 typhoon Tokage, and No. 0424 typhoon Nock-Ten. The second typhoon lasted longer, while the third one caused greater damage. The later, smaller bursts are their consequences. Since the focus is *typhoon*, most of the dates are negative. However, even though a day is "negative" in the tracking plot, not all of the related documents need be negative. Users can click "Details" to view the headlines and opinion polarities of these documents. Figure 3 shows the Chinese-English headlines of related documents on October 26, 2004.

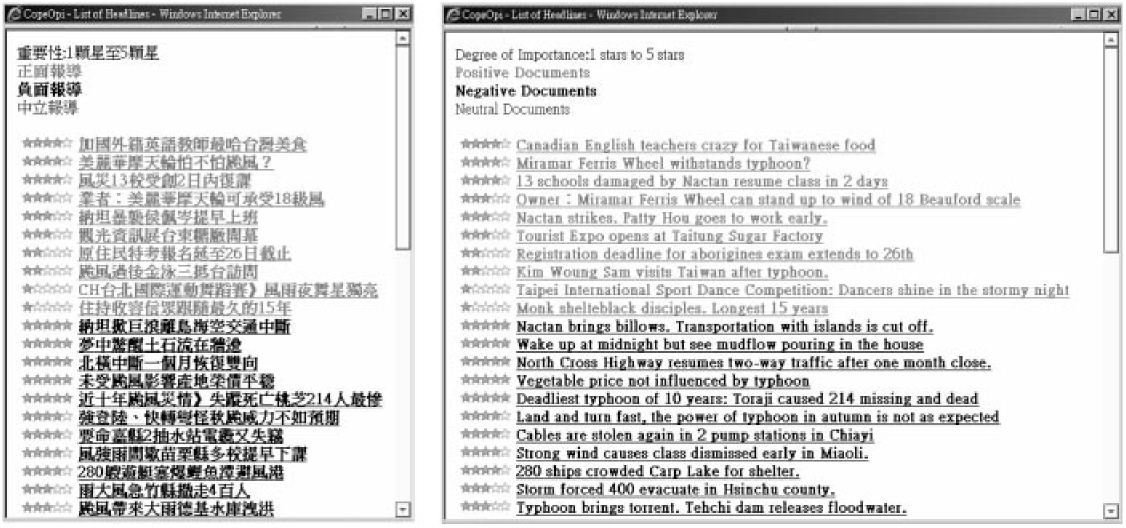
Of these 53 documents, 10 are positive and 43 are negative. The number of stars tells the degree of importance of each document. Five-star documents are judged the most informative among all of the documents for that day. Even though a given date is negative overall, there still may be positive documents for that day; for example, the second news article reports that "Miramar Ferris Wheel withstands typhoon." By clicking on the headline, CopeOpi displays the content of this document, as shown in Figure 4. Similarly, the bold gray sentences are positive, while the bold black ones are negative. Using the visualization in colored fonts (in the original system, green and red), users can easily

DOI: 10.1002/asi


CopeOpi - Tracking Plot - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
2004/10/18
2004/11/25
Details
完成
網際網路

Date : 20041026
Total : 53
Score : -124.97052967499996

FIG. 2. Typhoon tracking plot.


CopeOpi - List of Headlines - Windows Internet Explorer

重要性:1顆星至5顆星
正面報導
負面報導
中立報導

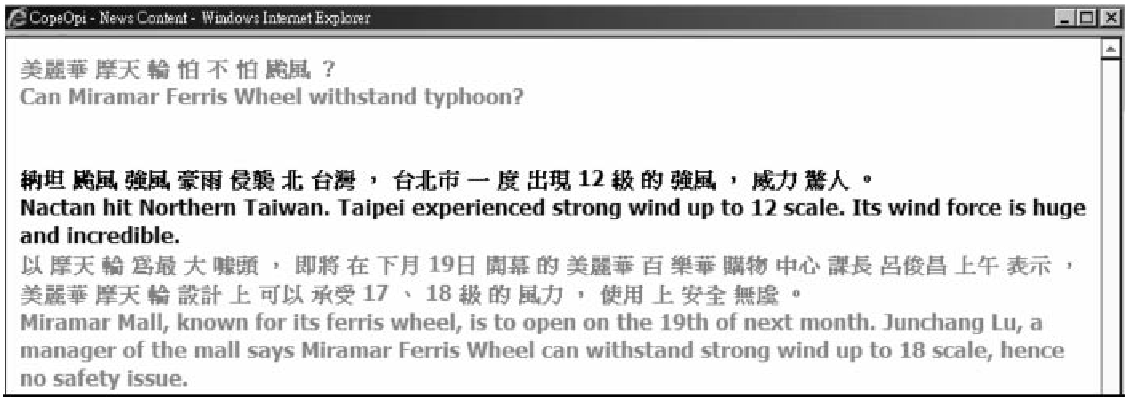
★★★★☆ 加國外籍英語教師最哈台灣美食
★★★★☆ 美麗華摩天輪怕不怕颱風？
★★★★☆ 風災13校受創2日內復課
★★★☆☆ 業者：美麗華摩天輪可承受18級風
★★★☆☆ 納坦暴襲侯佩岑提早上班
★★★☆☆ 觀光資訊展台東糖廠開幕
★★☆☆☆ 原住民特考報名延至26日截止
★★☆☆☆ 颱風過後金泳三抵台訪問
★☆☆☆☆ CH台北國際運動舞蹈賽》風雨夜舞星獨亮
★☆☆☆☆ 住持收容信眾跟隨最久的15年
★★★★★ 納坦掀巨浪離島海空交通中斷
★★★★★ 夢中驚醒土石流在牆邊
★★★★★ 北橫中斷一個月恢復雙向
★★★★★ 未受颱風影響產地菜價平穩
★★★★★ 近十年颱風災情》失蹤死亡桃芝214人最慘
★★★★☆ 強登陸、快轉彎怪秋颱威力不如預期
★★★★☆ 要命嘉縣2抽水站電纜又失竊
★★★★☆ 風強雨驟苗栗縣多校提早下課
★★★★☆ 280艘漁船塞爆鯉魚潭避風港
★★★☆☆ 雨大風急竹縣撤走4百人
★★★☆☆ 颱風帶來大雨德基水庫洩洪

CopeOpi - List of Headlines - Windows Internet Explorer

Degree of Importance:1 stars to 5 stars
Positive Documents
Negative Documents
Neutral Documents

★★★★☆ Canadian English teachers crazy for Taiwanese food
★★★★☆ Miramar Ferris Wheel withstands typhoon?
★★★★☆ 13 schools damaged by Nactan resume class in 2 days
★★★☆☆ Owner : Miramar Ferris Wheel can stand up to wind of 18 Beauford scale
★★★☆☆ Nactan strikes. Patty Hou goes to work early.
★★★☆☆ Tourist Expo opens at Taitung Sugar Factory
★★☆☆☆ Registration deadline for aborigines exam extends to 26th
★★☆☆☆ Kim Woung Sam visits Taiwan after typhoon.
★☆☆☆☆ Taipei International Sport Dance Competition: Dancers shine in the stormy night
★☆☆☆☆ Monk shelteblack disciples. Longest 15 years
★★★★★ Nactan brings billows. Transportation with islands is cut off.
★★★★★ Wake up at midnight but see mudflow pouring in the house
★★★★★ North Cross Highway resumes two-way traffic after one month close.
★★★★★ Vegetable price not influenced by typhoon
★★★★★ Deadliest typhoon of 10 years: Toraji caused 214 missing and dead
★★★★☆ Land and turn fast, the power of typhoon in autumn is not as expected
★★★★☆ Cables are stolen again in 2 pump stations in Chiayi
★★★★☆ Strong wind causes class dismissed early in Miaoli.
★★★★☆ 280 ships crowded Carp Lake for shelter.
★★★☆☆ Storm forced 400 evacuate in Hsinchu county.
★★★☆☆ Typhoon brings torrent. Tehchi dam releases floodwater.

FIG. 3. Headlines of opinion documents about typhoon "Nock-Ten."

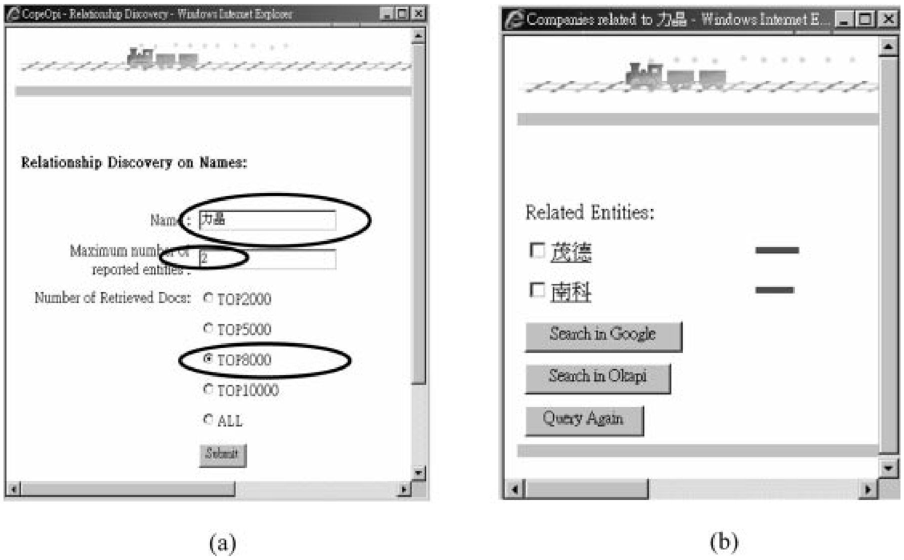


美麗華 摩天 輪 怕 不 怕 颱風 ？
Can Miramar Ferris Wheel withstand typhoon?

納坦 颱風 強風 豪雨 侵襲 北 台灣 ， 台北市 一 度 出現 12 級 的 強風 ， 威力 驚人 。
Nactan hit Northern Taiwan. Taipei experienced strong wind up to 12 scale. Its wind force is huge and incredible.
以 摩天 輪 爲最 大 噱頭 ， 即將 在 下月 19日 開幕 的 美麗華 百 樂華 購物 中心 課長 呂俊昌 上午 表示 ，
美麗華 摩天 輪 設計 上 可以 承受 17 、 18 級 的 風力 ， 使用 上 安全 無虞 。
Miramar Mall, known for its ferris wheel, is to open on the 19th of next month. Junchang Lu, a manager of the mall says Miramar Ferris Wheel can withstand strong wind up to 18 scale, hence no safety issue.

FIG. 4. Opinion sentences of a document that contains the target typhoon "Nock-Ten."

distinguish polarized sentences. In this example, the third sentence ("Miramar Mall . . . no safety issue") is clearly an opinion: that the Miramar Ferris wheel will be safe during the typhoon. Additionally, for our purposes we also consider the first and the second sentences to be opinions of the post author because they contain subjective information ("can . . . withstand typhoon," "wind force is huge and incredible").


CopeOpi - Relationship Discovery - Windows Internet Explorer
Submit
Companies related to 力晶 - Windows Internet E...
Search in Google
Search in Okapi
Query Again
(a)
(b)

Relationship Discovery on Names:

Name: 力晶

Maximum number of reported entities: 2

Number of Retrieved Docs: ○ TOP2000 ○ TOP5000 ● TOP8000 ○ TOP10000 ○ ALL

Related Entities:

□ 茂德

□ 南科

FIG. 5. An example of CopeOpi's relationship discovery.

Relationship Discovery

In terms of relationship discovery, CopeOpi finds entities related to a given target. Users input the named entities that they are interested in, specify the maximum number of entities to be returned, and choose the number of retrieved documents to be examined, for example, top 2000, 5000, 8000, 10000, or all. For instance, given the TOP2000 setting, the top 2000 documents among the retrieved documents are mined for further opinions. Figure 5(a) shows an example, where the analysis target is set to the company "力晶" (Powerchip Semiconductor Corp., or PSC), the number of the reported entities is set to 2, and the number of retrieved documents is set to the top 8,000 documents.

After discovering relationships, CopeOpi lists entities possessing strong relationships with the target, together with a measure of the strength of their relationship. The length of the horizontal line represents the strength of the relationship. The details of this operation will be described in following sections. Figure 5(b) shows two most related companies as discovered by CopeOpi, "茂德" (ProMOS Technologies) and "南科" (Nanya Technology Corp.). Two horizontal lines show that ProMOS has a stronger relationship with PSC than Nanya does. Indeed, these three companies are all DRAM manufacturers, and PSC and ProMOS are the two biggest DRAM companies in Taiwan.

Implementation of the CopeOpi System

Information retrieval, opinion extraction, opinion summarization, and opinion tracking are the four major modules for opinion mining. The retrieved documents must be relevant to a specific target; otherwise, the mined opinions would lack focus. We acquire the opinion words, sentences, and documents by opinion-extraction techniques. Opinion summaries (lists of headlines as in Figure 3) give an overview of the comments about the target, while summarized opinions (tracking plots as in Figure 2) are organized using temporal information. A summarized report, including the summarized opinion score and the number of retrieved documents, is given for each day (the right side of Figure 2). Detailed information is presented upon user request. The daily summarized opinion scores, arranged in a time sequence, form an opinion-tracking plot for the target.

Tracking plots display the changes of opinions towards one target over time. Comparisons can be made given the tracking plot of each target, which leads to multiple-target opinion tracking. Relationships among multiple targets are then discovered from their tracking plots by aligning time slots.

Opinion Mining

Opinion extraction provides useful information for mining opinions and relationships. The algorithms proposed by Ku and Chen (2007) are adopted in this paper. The basic idea is that the meaning of a Chinese word is a function of its composite Chinese characters. Likewise, the meaning of a Chinese sentence is a function of its composite Chinese words, and the meaning of a Chinese document is a function of its composite Chinese sentences. Therefore, we learn the distribution of characters in opinion words from training data and normalize it by Equations 1 and 2. The sentiment score of one character is defined as its normalized observation probabilities in positive opinion words minus its normalized observation probabilities in negative opinion words by Equation 3.

$$P_{c_i} = \frac{fp_{c_i} \Big/ \sum_{j=1}^{n} fp_{c_j}}{fp_{c_i} \Big/ \sum_{j=1}^{n} fp_{c_j} + fn_{c_i} \Big/ \sum_{j=1}^{n} fn_{c_j}}, \qquad (1)$$

$$N_{c_i} = \frac{fn_{c_i} \Big/ \sum_{j=1}^{m} fn_{c_j}}{fp_{c_i} \Big/ \sum_{j=1}^{n} fp_{c_j} + fn_{c_i} \Big/ \sum_{j=1}^{m} fn_{c_j}}, \tag{2}$$

$$S_{c_i} = (P_{c_i} - N_{c_i}), \tag{3}$$

where fp_{c_i} and fn_{c_i} denote the frequencies of a character c_i in the positive and negative words; P_{c_i} and N_{c_i} denote the weights of c_i as positive and negative characters, respectively; and n and m denote the total number of unique characters in positive and negative words, respectively.

We calculate the opinion score of one word according to the distribution probabilities of its composite characters in Equation 4. Here we use the bag-of-characters method.

$$S_w = \frac{1}{k} \times \sum_{j=1}^{k} S_{c_j}. \tag{4}$$

In Equation 4, k is the number of characters. By averaging the summation of the opinion score of each character c_j by k, we obtain the opinion score of the whole word S_w. The opinion score of one sentence is then calculated as a function of its composite words. Here the shallow sentence structure is considered due to the presence of negation words, which reverse the polarity of opinion words. The opinion score of one document is then simply the sum of the opinion scores of its composite sentences. Such a bottom-up methodology is suitable for mining information of different granularities. The opinions extracted at different levels are used in opinion summarization and tracking. However, merely summarizing opinions from documents does not give us sufficient information; we must further identify the events that engendered these opinions. Therefore, we track the opinions of one target first to generate the tracking plot in which we attempt to find temporal hints for the latent events.

Opinion tracking tells how people change their opinions over time. Tracking the opinions of a single target is fundamental for analyzing the variation of the target's reputation. Calculating the overall opinion scores of relevant documents for a specific target every day and display them by their temporal order generates a tracking plot. Whether a day is defined as *positive* or *negative* for a target is determined by the opinion tendency of that target for that day. Here positive and negative days are tracked separately to detect positive and negative events. Figure 6 shows an example of the tracking plot for Taiwan Semiconductor Manufacturing Company (TSMC). The black curve in the top figure illustrates the opinion scores on positive days.

In the tracking plot, the duration of positive periods is unknown. In other words, we have to determine the starting and ending date of every positive period to pinpoint the events in this period. However, if a positive curve drops and increases

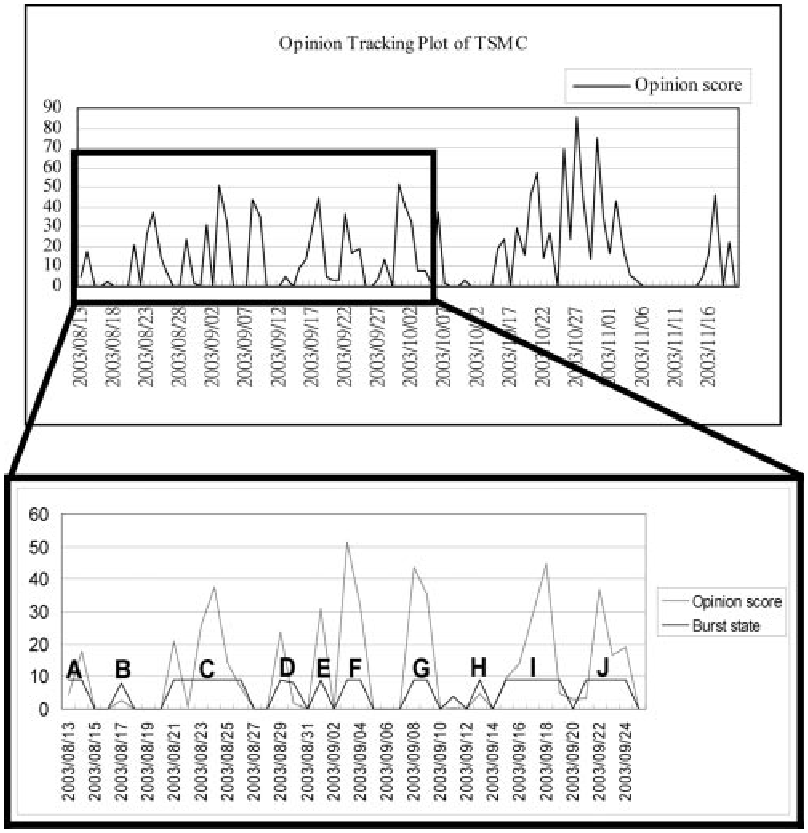


FIG. 6. Bursts detected from the TSMC positive opinion tracking plot.

DOI: 10.1002/asi

TABLE 1. Summaries of the event bursts in Figure 6.

Period	# of docs	From	To	Summary
A	47	2003/8/13	2003/8/15	TSMC and UMC expect revenue to rise in Q4.
B	36	2003/8/18	2003/8/18	Hedging through Morgan Taiwan stock index futures, foreign investors earn in both spot and futures markets.
C	113	2003/8/20	2003/8/26	Nasdaq hit record high of last 16 months, benefiting the market for long-term tradings.
D	21	2003/8/29	2003/8/30	IDM capacity is expected to grow low. Deluged with urgent analogic and IC orders, TSMC and UMC's low end plants receive orders over 10%.
E	22	2003/9/1	2003/9/1	Foreign capital keeps investing in TSMC and UMC.
F	52	2003/9/3	2003/9/4	Win-win to TSMC and ASE due to professional labor division.
G	68	2003/9/8	2003/9/9	0.18 um communication chip orders acquired by Chartered, Q3 capacity utilization exceeds 60%.
H	9	2003/9/13	2003/9/13	IBM wins another foundry order of power IC from Intersil.
I	39	2003/9/15	2003/9/19	3C electronics orders flooded in. TSMC and ASE capacity utilization approaches 95% in Q4.
J	81	2003/9/21	2003/9/24	Morris Chang: Semiconductor industry booming next year, and killer apps contribute.

again, it can be hard to know using only visualization if there are two events, or just one. To detect the proper period of events, we adopt a burst-detection approach (Kleinberg, 2002), which models a stream with a state automaton, where bursts appear as state transitions; the black curve in the bottom half of Figure 6 shows the resulting burst-detection plot. We thus generate opinion summaries based on documents within the same period, and find the events that resulted in these opinions. In Figure 6, symbols A–J denote the detected events. Take event A as an example: it ranges from the 13th to the 15th of August 2003. The brief summary generated is "TSMC and UMC expect revenue to rise in Q4." Table 1 shows the periods for event bursts A to J in Figure 6 as well as their summaries. We use the algorithms proposed by Ku (Ku et al., 2005) to decide how important a sentence is, and whether it is relevant and critical enough to be selected in summaries. When many sentences in one document are selected in summaries, this document is considered to be more important than others.

Events and opinions roughly correspond to causes and consequences. The same causes result in similar consequences for targets of a certain relationship. For two related targets (e.g., they are the same business, or they behave similarly in the stock market) the occurrence of an event will result in similar behavior, as represented in their opinion-tracking plots. This effect is illustrated in Figure 7.

The five curves in Figure 7 are tracking plots of the companies Taiwan Semiconductor Manufacturing Company (TSMC), United Microelectronics Corporation (UMC), Lucky Cement, a comparison of TSMC and UMC, and a comparison of TSMC and Lucky Cement, respectively. The TSMC and UMC plots in Figure 7(d) are more similar than those of TSMC and Lucky Cement (e). In fact, both TSMC and UMC are semiconductor companies, while Lucky Cement sells cement. We therefore assume that if the opinion-tracking plots of two companies are alike, they are more closely related than those with dissimilar plots. In the next section we will discuss relationship discovery using opinion-tracking plots, and compare this approach with conventional collocation-based approaches.

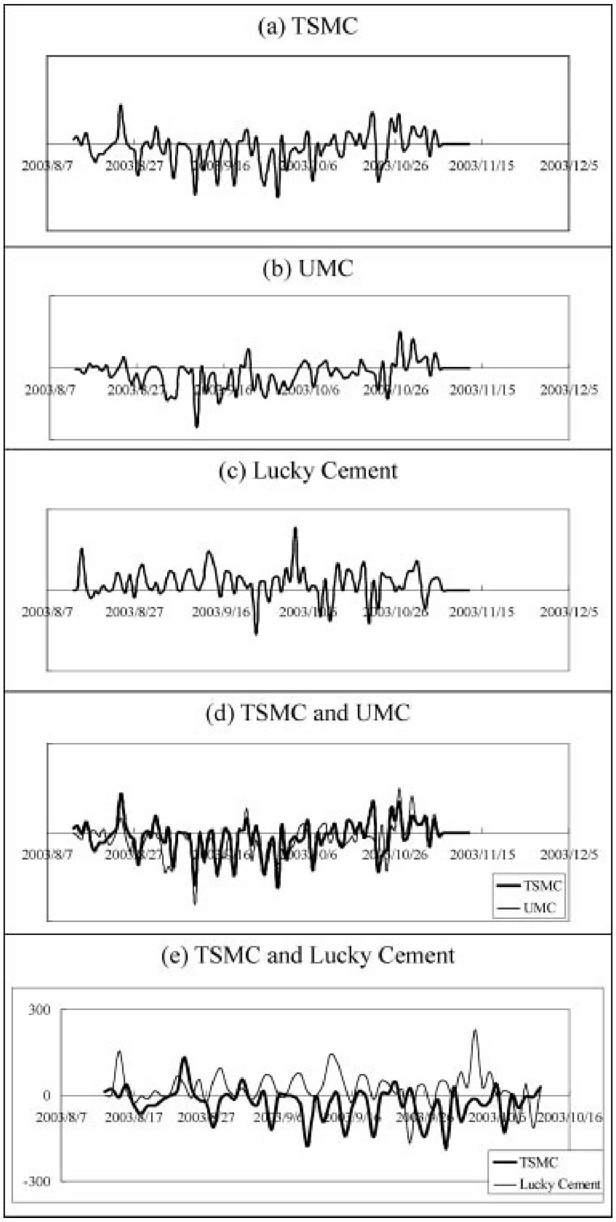


FIG. 7. Opinion Tracking for TSMC, UMC and Lucky Cement.

Relationship Discovery

Relationship discovery identifies certain relationships between targets. Targets can be any kind of object, including people, companies, or products. In addition, relationships can be explored among homogeneous and heterogeneous entities, such as persons and events. We propose two collocation-based models and four opinion-based models.

Collocation-based models. Collocation-based models discover relationships between two objects based on their in-context co-occurrences. Many statistical methods have been proposed. In this paper, we use mutual information and *t*-test for our collocation-based models because they are common in previous research (Manning & Schutze, 1999). Equations 5 and 6 shown below define these two collocation-based models.

$$I(A,B) = \log_2 \frac{P(A,B)}{P(A)P(B)}, \tag{5}$$

where $P(A,B)$ is the co-occurrence probability of the two targets A and B, and $P(A)$ and $P(B)$ are the occurrence probabilities of A and B. All pairs whose scores $I(A, B)$ are greater than zero are considered possible related pairs. They are ranked according to their scores in descending order and the requested number of pairs is returned from the top of the list.

$$t = \frac{\overline{x} - \mu}{\sqrt{\frac{s^2}{N}}}, \tag{6}$$

where $\overline{x}$ is the sample mean, s^2 is the sample variance, N is the sample size, and μ is the distribution mean. The *t*-test confidence level is 0.005 (99.5%); the value of t is 2.576.

The degree of collocation is separated into three levels in both models: document, sentence, and word levels. For document-level collocation, we count the number of documents in which the two targets co-occur. Similarly, for sentence-level collocation, we count the number of sentences in which the two targets co-occur. How frequently the two targets appear adjacent to each other within documents defines word-level collocation. For example, if A and B are two targets for mining relationships, we count occurrences of both "A B" and "B A".

Opinion-based models. Opinion-based models discover relationships between two objects based on their tracking-plot similarities. Each (x, y) point in an object's tracking plot shows that the object's summarized opinion score at date x is y, and as we have mentioned, this object's daily summarized opinion score comes from the sum of the various opinion scores of documents relevant to this object. The relationship strength is proportional to the overlap ratio of the two plots. We propose curve overlap (CO), digitalized curve overlap (DCO), and curve overlap with burst detection (BDCO).

The curve overlap model is as follows:

$$CO(A,B) = \frac{\sum_{i=1}^{n} \left(\mathrm{sgn}(R_i \cdot S_i) \cdot \frac{\min(|R_i|,|S_i|)}{\max(|R_i|,|S_i|)} \right)}{n}, \tag{7}$$

where R_i and S_i are the opinion scores of targets A and B on a specific day i, respectively, and n is the number of days in the tracking period.

The digitalized curve overlap model is represented by

$$DCO(A,B) = \frac{\sum_{i=1}^{n} \left(\mathrm{sgn}(R_i \cdot S_i) \cdot \frac{\min(\mathrm{sgn}(R_i),\mathrm{sgn}(S_i))}{\max(\mathrm{sgn}(R_i),\mathrm{sgn}(S_i))} \right)}{n}. \tag{8}$$

Only the sign of the opinion score is used to calculate the curve overlap in DCO. That is, only the opinion polarities are considered: The opinion degrees are not taken into consideration in this model.

In the curve overlap with burst detection model, first, $BD(X, t, i)$ is defined as the burst detection state on day i according to the tracking plot for target X. Since positive opinions and negative opinions are processed separately in burst detection, variable t identifies the tendency of the analyzed plot. When t is 1, function BD returns states from the positive tracking plot; when t is -1, function BD returns states from the negative tracking plot. Returned states are used to calculate the curve overlap in BDCO.

$$\begin{aligned}
&R_{BDi} = k \cdot \max(BD(A,1,i),\ BD(A,-1,i)),\\
&S_{BDi} = k \cdot \max(BD(B,1,i),\ BD(B,-1,i)),\\
&k = \begin{cases} 1 & \text{for} \quad \max(BD(X,1,i),\ BD(X,-1,i)) = BD(X,1,i) \\ -1 & \phantom{\text{for}} \quad \max(BD(X,1,i),\ BD(X,-1,i)) = BD(X,-1,i) \end{cases}.\\
&X \in \{A,B\}
\end{aligned} \tag{9}$$

$$BDCO(A,B) = \frac{\sum_{i=1}^{n} \left(\mathrm{sgn}(R_{BDi} \cdot S_{BDi}) \cdot \frac{\min(|R_{BDi}|,|S_{BDi}|)}{\max(|R_{BDi}|,|S_{BDi}|)} \right)}{n}. \tag{10}$$

The plot of burst detection states is a smoothed tracking-plot curve (see the black curve in the bottom figure in Figure 6). In this model, relationships are discovered from the burst plot as opposed to the tracking plot.

We adopt chi-square formula in the fourth opinion-based algorithm. The chi-square formula is defined in Equation 11, where f_{ij}^{o} is the exact observed value of frequency f_{ij}, and f_{ij}^{e} is the expected value of f_{ij}:

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij}^{o} - f_{ij}^{e})^2}{f_{ij}^{e}}. \tag{11}$$

Daily opinion scores are extracted from the results of opinion tracking, and the opinion-score signs are used to analyze target relationships. An opinion score of 0 means there are no documents retrieved on that day. Therefore, as shown in Table 2, we use a chi-square contingency table with one degree of freedom. For example, in Table 2, f_{11} denotes the number of dates for which target A and B's opinion scores are both positive, while f_{12} denotes the number of dates for which target A's opinion score is negative but target B's is positive.

TABLE 2. Chi-square contingency table with one degree of freedom.

		Signs of target A opinion score	
Number of days		+	−
Signs of target B opinion score	+	f_{11}	f_{12}
	−	f_{21}	f_{22}

Experimental Setup and Evaluation for Opinion Mining

When evaluating opinion-information-processing schemes, due to the subjective nature of the data, it is challenging to prepare a reasonable set of answer keys. We adopted two experiment datasets, and annotated a total of 1,073 news documents for opinion-mining experiments.

Experiment Materials

For opinion extraction, we used the NTCIR corpus. NTCIR[1] is one of three famous IR forums. For NTCIR-2, Chen and Chen (2001) developed the CIRB010 test collection for Chinese information retrieval. This collection consists of 50 topics and 132,173 Chinese documents covering the years from 1999 to 2002. Each topic in CIRB010 is in the TREC style. We first used part of CIRB010 for opinion extraction, and then used the NTCIR-6 opinion corpus—a larger corpus in which documents are selected from CIRB020 and CIRB040 (Chen, 2002)—to compare performance with other models in NTCIR-6. The CIRB020 and CIRB040 collections are composed of 249,203 and 901,446 documents, respectively, from which we selected 843 documents for 32 opinion tasks for the NTCIR-6 opinion corpus. All of the documents we used are news articles.

Of the 50 topics in CIRB010, we chose six topics that are more related to opinions instead of facts for opinion-extraction experiments. We selected a total of 192 documents, which are judged as relevant by NTCIR assessors to these six topics. This corpus is denoted as CIRB010-OP hereafter. We went a step further by a larger scale experiment in the NTCIR-6 pilot task. In this task, a total of 843 documents, which are judged as relevant to 32 topics in CIRB020 and CIRB040 by NTCIR assessors, were used. This corpus is denoted as CIRB020/040-OP hereafter. The topics and the corresponding number of documents are shown in Table 3.

According to the opinion-extraction methods proposed above, we only need to estimate the probabilities of characters to calculate their opinion scores, and then we can calculate the opinion scores of words and sentences in terms of them. Therefore, positive and negative opinion words in the sentiment dictionary NTUSD[2] were used to calculate the observation probabilities of characters by Equations 1 and 2, and further to calculate the opinion scores of these characters. After that we calculated opinion scores of words and sentences. In our experiments, sentences were all used for testing since their opinion scores can be calculated without extra training processes.

Gold Standard Acquisition

The results of opinion annotation vary given different perspectives. Generally, multiple annotators are needed to prepare answers for opinion mining (Ku et al., 2006). Here all of the documents for the opinion-mining experiments were tagged by three annotators. Because sentences are the most reasonable units for expressing opinions, and because extracting opinion sentences has been shown to be the most challenging among all granularities (Ku & Chen, 2007), we focus on sentence-level opinion-mining evaluation in this paper. That is, we annotated sentences in both CIRB010-OP and CIRB020/040-OP, and the latter were adopted in the NTCIR-6 pilot task (Seki et al., 2007). There are two metrics for evaluation: Under the strict metric, only sentences for which the three annotations are all consistent are counted. Under the lenient metric, only sentences with two or three consistent annotations are counted; the majority annotation is treated as the gold standard.

Experiments and Results

To show the capability of the proposed algorithms, we evaluated sentence-level opinion extraction. Tables 4 and 5 show the performance of sentence-level opinion extraction on two different data sets, CIRB010-OP and CIRB020/040-OP. When experimenting with CIRB010-OP, we considered the relevance issue—that the returned sentences should be relevant and should be opinions. For the CIRB010-OP-1 setting, we assumed that the extracted opinion sentences were all relevant, whereas for the CIRB010-OP-2 setting, we adopted the method proposed by Ku et al. (2005) to determine the relevance. This method takes into consideration the distribution of terms in documents and paragraphs in order to select important and representative terms, and retrieves the sentences containing these terms as relevant to the topic of the whole document set. For the CIRB010-OP-3 setting, we used the concept words defined by the topic creators and selected sentences containing these words as relevant sentences. Table 4 shows that relevance is important when mining opinions, and also shows that manual keywords are better than automatic keywords, and using either kind of keywords is better than not using keywords. The CIRB010-OP-2 setting is only slightly better than CIRB010-OP-1 because sentence relevance retrieval is difficult (Soboroff & Harman, 2003; Ku et al., 2005).

Table 5 shows the performance using the larger corpus, CIRB020/040. For the CIRB020/040-OP-A setting, we retrieved relevant documents automatically by the method proposed by Ku et al. (2005), and then selected those with opinions, while for the CIRB020/040-OP-B setting we extracted opinions from the given relevant sentences, i.e.,

[1] http://research.nii.ac.jp/ntcir/

[2] http://nlg18.csie.ntu.edu.tw:8080/opinion/index.html

TABLE 3. Topics and number of documents in NTCIR corpora for opinion mining.

Topic ID	Topic title	# of docs
CIRB010-OP		
ZH021	Civil ID Card	37
ZH024	The Abolishment of Joint College Entrance Examination	55
ZH026	The Chinese-English Phonetic Transcription System	30
ZH027	Anti-Meinung Dam Construction	14
ZH028	Logging of Chinese Junipers in Chilan	23
ZH036	Surrogate Mother	33
	Total	192
CIRB020/040-OP		
001	Time Warner, American Online (AOL), Merger, Impact	13
002	President of Peru, Alberto Fujimori, scandal, bribe	28
003	Kim Dae Jun, Kim Jong Il, Inter-Korea Summit	13
004	US Secretary of Defense, William Sebastian Cohen, Beijing	13
005	G8 Okinawa Summit	38
006	Wen Ho Lee Case, classified information, national security	41
007	Ichiro, Rookie of the Year, Major League	14
008	Jennifer Capriati, tennis	25
009	EP-3 surveillance aircraft, F-8 fighter, aircraft collision	95
010	History Textbook Controversies, World War II	13
011	Tobacco business, accusation, compensation	19
012	Tiger Woods, sports star	11
013	"Qiudou" (Autumn Struggle), Appeal, Laborer, Protest, Taiwan	28
014	Expert, Opinion, International Monetary Fund (IMF), Asian countries	21
015	Find articles dealing with a teenage social problem	19
016	Divorce, Family Discord, Criticisms	24
017	China, Reaction, Taiwan, Diplomatic Relations	14
018	China, Stationing, Weapons, Taiwan	17
019	Animal Cloning Technique	21
020	Sexual Harassment, Lawsuits	55
021	Olympic, Bribe, Suspicion	21
022	North Korea, Daepodong, Asia, Response	127
023	Joining WTO	89
024	China Airlines Crash	13
025	Province-refining	14
026	Economic influence of the European monetary union	32
027	President Kim Dae-Jung's policy toward Asia	12
028	Clinton scandals	71
029	War crimes lawsuits	23
030	Nuclear power protests	13
031	College Admission Policy	28
032	Youth Counseling	13
	Total	843

TABLE 4. Opinion extraction performance for CIRB010-OP.

	CIRB010-OP-1	CIRB010-OP-2	CIRB010-OP-3
Precision	0.341	0.381	0.578
Recall	0.681	0.648	0.672
f-measure	0.454	0.480	0.622

TABLE 5. Opinion extraction performance for CIRB020/040-OP.

	CIRB020/040-OP-A	CIRB020/040-OP-B
Precision	0.335	0.664
Recall	0.448	0.890
f-measure	0.383	0.761

the sentences judged as relevant by the NTCIR assessors as mentioned above. Clearly, CIRB020/040-OP-B is much better than CIRB020/040-OP-A.

Tables 6 and 7 show the performance of all of the systems in the NTCIR-6 pilot task (Seki et al., 2007). All employ the CIRB020/040-OP corpus under the lenient metric in their evaluations. For opinion-sentence extraction, our system is in the high performance group. For opinion- and relevant-sentence extraction, we ranked the second.

Experimental Setup and Evaluation for Relationship Discovery

We collected 1,282,050 economics-related documents as the knowledge base for relationship discovery. We used

TABLE 6. Opinion sentence extraction in the NTCIR-6 pilot task.

	Opinion		
Group	P	R	F
UMCP-1	0.645	0.974	0.776
UMCP-2	0.630	0.984	0.768
Gate-1	0.643	0.933	0.762
NTU	0.664	0.890	0.761
Gate-2	0.746	0.591	0.659
CHUK	0.818	0.519	0.635
ISCAS	0.590	0.664	0.625

TABLE 7. Opinion and relevant sentence extraction in the NTCIR-6 pilot task.

	Opinion and polarity		
Group	P	R	F
CHUK	0.522	0.331	0.405
NTU	0.335	0.448	0.383
UMCP-1	0.292	0.441	0.351
UMCP-2	0.286	0.446	0.348
ISCAS	0.232	0.261	0.246
Gate-1	—	—	—
Gate-2	—	—	—

1,078 listed companies in Taiwan as the target candidates for discovering relationships, as well as the stock indices of these companies during the time period represented in the knowledge base to verify the correctness of the discovered relations.

Experiment Materials

For model training, a total of 1,282,050 economics-related documents were collected automatically from 93 Web sources from August 2003 to May 2005. For collocation-based models, we used all of the documents to count co-occurrences. For opinion-based models, we first retrieved documents mentioning 1,078 listed companies using an Okapi IR system (Robertson, Walker, & Beaulieu, 1998) from these economics-related documents. Because the opinion analyses require sufficient information, we sorted the 1,078 companies by decreasing number of the retrieved documents for each company and selected the top 250 as the relationship-discovery targets. Here a retrieved document for a target means a document retrieved by an IR system using the target as the query. A query for a company is the company name and its stock-market abbreviation. If a document mentions the company name or its abbreviation, it will be retrieved and viewed as relevant. In reality, a mere mention does not always indicate relevance, but still provides some information about the target. We ranked in descending order the retrieved documents by their relevance to the query calculated by the IR system, and then prepared for our experiments five document sets, i.e., the top 2,000, 5,000, 8,000, 10,000, and all, for each company. On average, we retrieved 10,441 documents for each company.

Gold Standard Acquisition

For relationship discovery, it is difficult to find the correct answers because there are many kinds of relationships of different strengths. In this paper, the real-world behavior of the listed companies was helpful for generating answers. Companies whose stock prices behaved similarly in the market over a fairly long period of time were considered related, regardless of whether these companies were competitors, had the same investors, or belonged to the same vertical market.

Generally speaking, investors want to know the relationships among designated companies in the stock market, which makes relationship discovery a practical application. To match real-world phenomena, we mined the gold standards from the stock prices from August 2003 to May 2005, i.e., the same period as the corpus publication time. We postulated that two stocks that always rise and fall together are correlated, and adopted the chi-square method to find such pairs from a total of 31,125 (C_2^{250}) company pairs. The change of the stock price of one company is defined in Equation 12. To be objective, we took into account the large-cap price change to decide the ups (↑) and downs (↓) of the stock.

$$d_i = \frac{p_i - p_{i-1}}{p_{i-1}} - \frac{q_i - q_{i-1}}{q_{i-1}}, \tag{12}$$

where p_i (p_{i-1}) is the price of the stock and q_i (q_{i-1}) is the price of the large-cap on day i ($i - 1$), and d_i is the percent of the stock price difference. If d_i is positive, an "up" (↑) appears on day i; if d_i is negative, a "down" (↓) appears for that day.

We adopted the chi-square formula, as defined in Equation 11. Taking into account the ups (↑) and downs (↓) of the stock prices of two companies *A* and *B*, the chi-square contingency table with one degree of freedom is generated in Table 8. When taking unchanged prices into consideration, the chi-square table with four degrees of freedom can also be generated in Table 9.

With different degrees of freedom and significance levels, we extracted the correlated company pairs as the gold standard. Table 10 shows the number of pairs in the gold standard for different conditions.

In this case, we decided that specific company pairs with strong relationships were more informative than a large

TABLE 8. Chi-square contingency table with one degree of freedom.

		Stock price of company *A*	
Number of days		↑	↓
Stock price of company *B*	↑	f_{11}	f_{12}
	↓	f_{21}	f_{22}

TABLE 9. Chi-square contingency table with four degrees of freedom.

		Stock price of company *A*		
Number of days		↑	–	↓
Stock price of company *B*	↑	f_{11}	f_{12}	f_{13}
	–	f_{21}	f_{22}	f_{23}
	↓	f_{31}	f_{32}	f_{33}

TABLE 10. Numbers of company pairs.

Degree of freedom	$\chi^2_{.950}$	$\chi^2_{.990}$	$\chi^2_{.995}$
1	7,815	4,239	2,008
4	2,489	1,366	703

number of company pairs with weak relationships; that is, that precision was more important than recall. Therefore, we used the strictest condition to generate the gold standard, and selected a total of 703 pairs as the gold standard under $\chi^2_{.995}$ with four degrees of freedom.

Experiments and Results

Since the performance of companies is reflected in their stock prices, and because stock prices are public information, we chose companies as the entity type when evaluating methods for relationship discovery.

Relationship discovery performance. The relevance issue is also important in relationship discovery evaluation. Only by working on relevant documents can we gauge the real performance of both collocation-based and opinion-based models. Otherwise we cannot decide whether these models perform well, because extracting relationships from documents irrelevant to the target may result in too much noise. However, retrieved documents are only those documents that mention the targets (companies). Notably, many companies adopt auspicious or general terms such as *happy*, *lucky*, and *peace* as part of their name, and these terms are often used as the abbreviations of these companies. We found that attempting to retrieve documents relevant to these targets by submitting the complete company name such as "lucky cement" as a phrase to the IR system will miss lots of relevant documents, but retrieving documents by submitting only these terms (e.g., *lucky*) yielded many irrelevant documents. To examine the influence of relevance on relationship discovery, we also performed experiments excluding companies like these, and a total of 53 companies were excluded here.

As mentioned above, we used 703 pairs to evaluate relationship discovery. We focused on whether there was some relationship between the two companies in each pair proposed by our methods. Therefore, we use the precision metric to compare the results of relationship discovery, but also list recall, *f*-measure, and significance of results for reference.

TABLE 11. The MI model in relationship discovery including companies with general names.

MI	Document level			Sentence level			Word level		
N	P	R	F	P	R	F	P	R	F
25	0.24	0.01	0.02	**0.96*****	0.03	0.07	0.48	0.02	0.03
50	0.26	0.02	0.04	**0.94*****	0.07	0.13	0.46	0.03	0.06
100	0.25	0.04	0.06	**0.74*****	0.11	0.18	0.44	0.06	0.11
200	0.24	0.07	0.10	0.54***	0.15	0.24	0.40	0.11	0.18
500	0.18	0.13	0.15	0.32†	0.23	0.27	0.30	0.21	0.25

$^{***}p \le 0.005$. $^{\dagger}p > 0.1$.

TABLE 12. The *t*-test model in relationship discovery including companies with general names.

t-test	Document level			Sentence level			Word level		
N	P	R	F	P	R	F	P	R	F
25	0.44	0.02	0.03	0.48	0.02	0.03	0.60†	0.02	0.04
50	0.26	0.02	0.04	0.46	0.03	0.06	0.46†	0.03	0.06
100	0.24	0.03	0.06	0.44	0.06	0.11	0.41†	0.06	0.10
200	0.21	0.06	0.09	0.34	0.10	0.15	0.37†	0.08	0.13
500	0.15	0.11	0.12	0.25	0.18	0.21	0.34***	0.10	0.15

$^{***}p \le 0.005$. $^{\dagger}p > 0.1$.

TABLE 13. The MI model in relationship discovery excluding companies with general names.

MI	Document level			Sentence level			Word level		
N	P	R	F	P	R	F	P	R	F
25	0.28	0.01	0.02	**0.96*****	0.04	0.08	0.52	0.02	0.04
50	0.28	0.02	0.04	**0.96*****	0.08	0.15	0.56	0.05	0.09
100	0.25	0.04	0.07	**0.75*****	0.12	0.21	0.49	0.08	0.14
200	0.24	0.08	0.12	0.54‡	0.18	0.27	0.44	0.14	0.22
500	0.18	0.15	0.17	0.32†	0.26	0.29	0.32	0.26	0.29

$^{***}p \le 0.005$. $^{\dagger}p > 0.1$. $^{\ddagger}p \le 0.1$.

MI and t*-test.* Tables 11 and 12 show the results of MI and *t*-test. MI performs better than the *t*-test, and MI achieves precision rates of 0.96, 0.94, and 0.74 when the top 25, 50, and 100 answers are proposed, respectively. Tables 13 and 14 show the results of MI and *t*-test when companies whose names are general terms are excluded; there is a slight improvement. The best precision rates of MI rise up to 0.96, 0.96, and 0.75. Results of sentence-level MI are significantly better than those of document-level and word-level MI. In contrast, although the word-level *t*-test performs better than the document-level and sentence-level *t*-tests, the improvement is not very significant.

CO, DCO, BDCO, and χ^2*.* Table 15 shows the average performance of opinion-based models using different numbers of retrieved documents. The top 2,000, 5,000, 8,000, 10,000, and all of the retrieved documents of the two companies

TABLE 14. The *t*-test model in relationship discovery excluding companies with general names.

t-test	Document level			Sentence level			Word level		
N	P	R	F	P	R	F	P	R	F
25	0.44	0.02	0.04	0.48	0.02	0.04	$0.64^{\dagger}$	0.03	0.05
50	0.26	0.02	0.04	0.44	0.04	0.07	$0.54^{\dagger}$	0.05	0.08
100	0.26	0.04	0.07	0.43	0.07	0.12	$0.47^{\dagger}$	0.08	0.13
200	0.23	0.08	0.11	0.34	0.11	0.17	$0.41^{\dagger}$	0.10	0.16
500	0.16	0.13	0.14	0.26	0.21	0.23	0.42^{***}	0.14	0.21

$^{***}p \le 0.005$. $^{\dagger}p > 0.1$.

are used for relationship discovery. Retrieving the top 8,000 retrieved documents is the best strategy for CO, DCO, BDCO, and χ^2. Insufficient retrieved documents (top 2,000, 5,000) as well as noise (top 10,000, all) result in deteriorated performance. Table 16 shows the comparison of four opinion-based models using top 8,000 retrieved documents, and Table 17 shows the performance of the same four models when companies with general names are filtered out. In opinion-based models, the effect of filtering out general terms improves performance much more than in collocation-based models. Both selecting the proper number of documents and filtering out general terms eliminates many nonrelevant documents. This shows that opinion-based models are more sensitive to relevance than collocation-based models, except for the chi-square model. In this case, the chi-square model is affected more by the distribution of ups and downs than by the curve shape. Therefore, the chi-square model is less sensitive to the difference of curves than the other three models.

Different opinion-based models are compared in Table 17. CO is the best among all the models; it performs significantly better than DCO and BDCO, and it achieves precision rates of 0.92, 0.86, and 0.68 when the top 25, 50, and 100 answers are proposed, respectively. DCO, which utilizes the digitized tracking plot, and BDCO, which discovers relationships from the smoothened tracking plot, perform worse than CO. We can conclude that the opinion weights (CO vs. DCO) and the changes in a short period (CO vs. BDCO) are both important clues for relationship discovery. χ^2 achieves precision comparable with CO; there is no significant difference

TABLE 15. Opinion-based models using different numbers of retrieved documents.

	Top 2,000			Top 5,000			Top 8,000			Top 10,000			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
CO	0.472	0.078	0.117	0.552	0.092	0.137	0.603	0.102	0.152	0.606	0.104	0.155	0.510	0.091	0.135
DCO	0.223	0.034	0.055	0.334	0.051	0.082	0.395	0.055	0.090	0.363	0.053	0.086	0.304	0.049	0.079
BDCO	0.090	0.012	0.020	0.049	0.006	0.009	0.097	0.014	0.023	0.092	0.014	0.023	0.107	0.018	0.029
χ^2	0.550	0.069	0.114	0.584	0.072	0.118	0.621	0.078	0.128	0.625	0.079	0.130	0.600	0.077	0.126

TABLE 16. Opinion-based models in relationship discovery including companies with general names.

Top 8,000	CO			DCO			BDCO			χ^2		
N	P	R	F	P	R	F	P	R	F	P	R	F
25	0.840^{**}	0.030	0.058	0.440	0.016	0.030	0.120	0.004	0.008	0.960	0.034	0.066
50	0.780^{*}	0.056	0.104	0.540	0.038	0.072	0.100	0.007	0.013	0.700	0.050	0.093
100	0.640^{***}	0.091	0.159	0.420	0.060	0.105	0.110	0.016	0.027	0.530	0.075	0.132
200	0.475^{***}	0.135	0.210	0.295	0.084	0.131	0.085	0.024	0.038	0.400	0.114	0.177
500	0.278^{*}	0.198	0.231	0.210	0.149	0.175	0.050	0.036	0.042	0.232	0.165	0.193

$^{*}p \le 0.05$. $^{**}p \le 0.01$. $^{***}p \le 0.005$.

TABLE 17. Opinion-based models in relationship discovery excluding companies with general names.

Top 8,000	CO			DCO			BDCO			χ^2		
N	P	R	F	P	R	F	P	R	F	P	R	F
25	0.920^{*}	0.038	0.073	0.640	0.026	0.051	0.680	0.028	0.054	0.960	0.040	0.076
50	0.860^{*}	0.071	0.131	0.660	0.054	0.101	0.640	0.053	0.097	0.700	0.058	0.107
100	0.680^{***}	0.112	0.192	0.460	0.076	0.130	0.460	0.076	0.130	0.550	0.091	0.156
200	0.480^{**}	0.158	0.238	0.330	0.109	0.164	0.345	0.114	0.171	0.400	0.132	0.198
500	0.276^{*}	0.227	0.249	0.216	0.178	0.195	0.224	0.185	0.202	0.236	0.194	0.213

$^{*}p \le 0.05$. $^{**}p \le 0.01$. $^{***}p \le 0.005$.

TABLE 18. Intersection and difference.

Top N	MI ∩ CO	MI-CO	CO-MI
25	16	8	7
50	27	21	16
100	43	32	25
200	67	40	29
500	103	57	35

TABLE 19. Average rank of CO-MI and MI-CO.

Top N	Average rank of CO-MI in MI	Average rank of MI-CO in CO
25	614.43	180.25
50	806.75	439.29
100	1,305.12	722.03
200	1,487.86	1,085.88
500	2,487.69	3,663.12

between their performance. However, the χ^2 precision rate drops rapidly when more answers are proposed. Therefore, CO is the best of the opinion-based models.

We also examine the proposed answers of the collocation-based and opinion-based models. The answers of the two best models, the collocation model MI and the opinion-based model CO, are selected as the targets for analysis. Table 18 shows the intersection and difference of the two answer sets: Only about half of the proposed answers of CO and MI intersect, and this quantity decreases when more answers are proposed. This result shows that MI and CO propose different answers. Table 19 further shows the ranks of answers in the set differences. The set difference CO-MI consists of answer pairs proposed by CO but excludes answer pairs in the top *N* of MI as well as pairs that were not proposed by the MI model. Similarly, the set difference MI-CO consists of answer pairs proposed by MI but excludes answer pairs in the top *N* of CO as well as pairs that were not included by the CO model. For example, if CO proposed the company pairs {1,2,7,5,3,8} and MI proposed the company pairs {1,2,4,6,8,3,9}, in the top 5 case, CO equals {1,2,7,5,3}, MI equals {1,2,4,6,8} and MI ∩ CO equals {1,2}. CO-MI equals {3} because MI also finds company pairs {1,2}, and {7,5} are excluded by MI. Likewise, MI-CO equals {8}. All of the answers in CO-MI are checked to see which ranks they are in the answer set of MI; likewise for the MI-CO set. If the ranks of the answers found by one model are low in the other model, it means that the other model does not propose the answers found by this model. Table 19 shows that the MI ranks of the answers in CO-MI tend to be lower than the CO ranks in MI-CO. In other words, the CO model finds company pairs that do not often co-occur and are thus not found by the MI model.

To illustrate the idea that the CO and MI models complement each other, we further analyzed these company pairs and identified conditions in which related company pairs were found by the CO model but not by MI. They are listed as follows:

1. The company pair was never listed together in a single document in the news corpus, or when the two companies did co-occur in a document, it was with too many other companies, weakening the impact for the MI model. In news articles, companies in the same industry are usually grouped for discussion. For example, Hung Sheng Construction Co. (2534.TW) is in the Building Materials and Construction group and China Man-Made Fiber Corporation (1718.TW) in the Chemical group. They are seldom mentioned in the same documents. Even in the few documents in which they both are mentioned, these documents are overviews where many other companies are also mentioned. Generally, news articles seldom bring to attention relationships between materials (China Man-Made Fiber Corporation) and manufacturers (Hung Sheng Construction Co.), but instead focus more on the development of individual industry groups. However, such companies do influence each other in the vertical market. Therefore, the CO model is able to discover their relationship by analyzing changes of opinions about the two companies.
2. The relationship of the company pair is not direct, or these two companies can only be connected by keywords or concepts. For example, Chien Hsing Stainless Steel (2025.TW) and Hsin Kuang Steel Co. (2031.TW) are both in the Iron and Steel group. However, in the news they seem to be selected as a representative for this industrial group only randomly, because their company sizes are comparable. Therefore, they did not often co-occur in news articles. Apparently, their industrial group is their only link; their relationship was not easily observed directly from the news context.
3. The sizes of the two companies are not comparable, or the numbers of extracted documents for mining are not comparable. For example, the related pair AU Optronics Corp. (2409.TW) and QDI (3012.TW) in the optoelectronic industry was not extracted by the MI model. This is because AU Optronics Corp. is much larger than QDI and appeared in an overwhelming number of articles as compared to QDI. The MI score of this pair is low, though if we only check pairs containing AU Optronics Corp., the pair AU Optronics Corp. and QDI still had the highest MI score. Sometimes, even though the sizes of two companies are comparable, if there is an event only related to one company, for example, a CEO scandal, the numbers of documents for these two companies differ greatly, resulting in a decreased MI score.

Given the above analysis, we know that opinions on two targets, or opinions on a single target and an event that involves other targets, are helpful for relationship discovery. The CO model relies on opinions and events to link targets. Opinions on one target and events involving only this target do not aid relationship discovery, because these opinions have no direct or indirect influence on other targets.

Integration models. Given that the answer sets of the best model of two types, i.e., CO and MI, are complementary, we evaluated two integration models to see whether taking

TABLE 20. CO + MI and CO ∩ MI without general names.

	CO + MI			CO ∩ MI		
N	P	R	F	P	R	F
25	$1.00^{††}$	0.04	0.08	1.00	0.04	0.08
50	$0.92^{††}$	0.08	0.14	0.92	0.08	0.14
100	$0.79^{††}$	0.13	0.22	0.76	0.13	0.21
200	$0.64^{***‡}$	0.21	0.31	0.62	0.20	0.30
500	$0.37^{***‡}$	0.31	0.34	0.39	0.32	0.35

$^{††}p > 0.1$. $^{‡}p \leq 0.1$. $^{***}p \leq 0.005$.

into account both opinions and collocations would benefit relationship discovery. The CO + MI model considers both CO and MI scores, while the CO ∩ MI model returns answers present in both the CO and MI answer sets. The formula for the CO + MI model is defined as follows.

CO + MI.

$$CO + MI(A,B) = \alpha\frac{MI(A,B)}{\rho} + \beta\frac{CO(A,B)}{\upsilon}. \qquad (13)$$

where $\alpha = 0.5$, $\beta = 0.5$, and ρ and υ are normalization constants.

CO ∩ MI. CO ∩ MI scans the MI and CO answers in a round-robin fashion to select common candidates. Thus, CO + MI integrates two types of information using scores, while CO ∩ MI integrates based on ranks.

Compared with the collocation-only and opinion-only models, both integration models perform better (results of two significance tests are shown in the P column). Table 20 shows the performance of CO + MI and CO ∩ MI. CO + MI achieves precision rates of 1, 0.92, and 0.79, and CO ∩ MI achieves precision rates of 1, 0.92, and 0.76, respectively, when the top 25, 50, and 100 answers are proposed. The overall performance of the eight models in Figure 8 shows that CO + MI is the best of all the models. The fact that CO + MI slightly outperforms CO ∩ MI tells us indirectly that CO and MI are both good algorithms for selecting correct related pairs (precision-based), as opposed to only collecting potential candidates (recall-based). The symbols shown after the precision rates of CO + MI denote the significant levels when comparing CO + MI with CO and MI, respectively. The CO + MI results improve significantly with the number of reported related pairs.

Mined Relationships

We attempted to mine objects that led to opinions of the same polarities during a given time period, and extracted them with the assumption that they represented some kind of relationship. People may be curious about these relationships: Can the types of possible relationships among these objects be prior knowledge? Can we find these relationship types in other information sources? Here these predefined relationships denote commonly accepted relationships; for companies, they would include ownership, partnership, competitors, and so forth.

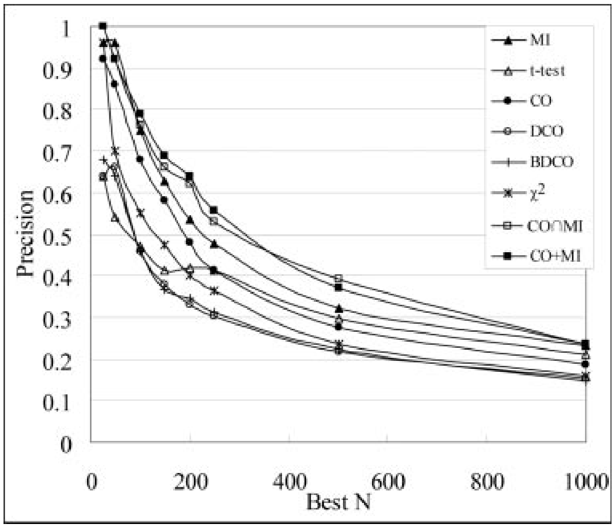


FIG. 8. Performances of all models.

In this section, we use company pairs to demonstrate our idea, because a company's performance can be summarized by its stock indices, and because it facilitates the evaluation process. We take a closer look at the relationships of company pairs extracted as the gold answers; they are listed in Table 21 (the top 20 company pairs, ranked by the strength of relationship between each pair). The pairs are composed of the companies Com:A and Com:B, and the industry types are those defined by the Taiwan Stock Exchange Center.

Table 21 shows that 14 company pairs among the top 20 are in the same industry. Among the other 6 company pairs (in bold fonts) are other types of relationships. Some of these pairs produce similar but not identical products, and some of them are in the same vertical market. Thus, we found that not all objects with similar performance are in the same industries, that is, have the same type of relationship. In contrast, not all companies with the same type of relationship (e.g., the same industry) have similar performance. For example, not all banks are extracted in the gold answer set. We may therefore say that the predefined relationships are different from the relationships extracted according to opinions on objects or according to their performance, and that we cannot find the predefined relationships from the performance-based ones easily, and vice versa.

The characteristics of performance-based relationships are also distinct from those of predefined ones. Predefined relationships give users information that may be acquired from other sources. Performance-based relationships show users which objects behave similarly, i.e., perform similarly, thus yielding similar opinions. As mentioned earlier, performance and opinions are highly related to events. Therefore, relationships mined from performance and opinions provide

DOI: 10.1002/asi

TABLE 21. Top 20 company pairs and their industry types.

Rank	Com:A	Com:B	Industrial group	Rank	Com:A	Com:B	Industrial group
01	CGPC	USIFE	**PVC/PE**	11	HuaKu	HHE	Building
02	CPTT	QDI	TFT-LCD	12	AUO	CMO	TFT-LDC
03	Yeunchyang	Tachen	Stainless Steel	13	Promos	NTC	DRAM
04	PSC	Promos	DRAM	14	CHB	TBB	Bank
05	CPTT	Hannstar	TFT-LCD	15	YiehPhui	ChSteel	**SGCC/SPHC**
06	Everlight	Epistar	**LED Packing/LED Grain**	16	GPPC	USIFE	**SM/PE**
07	PSC	NTC	DRAM	17	UPC	USIFE	**DOP/PE**
08	Firstholding	HNFHC	Bank	18	HSC	Kindom	Building
09	Tachen	CSSSC	Stainless Steel	19	THS	FengHsin	**Structural steel/Steel bar**
10	Yeunchyang	CSSSC	Stainless Steel	20	TBB	TCB	Bank

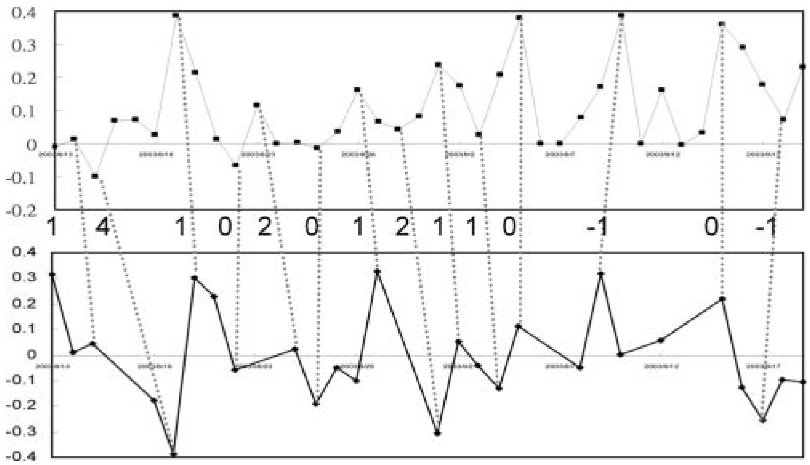


FIG. 9. Tracking plot vs. Stock chart (BenQ).

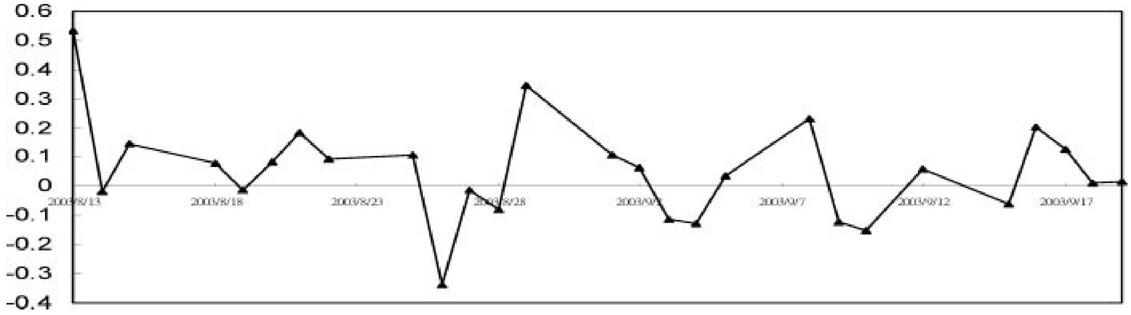


FIG. 10. Stock chart of the large cap.

information about which objects will also be influenced when events happen to another object.

Opinion-Based Prediction

The last section deals with how to discover long-term relationships from opinions. Extracting correlated company pairs from tracking plots performs well in comparison to the gold standard extracted from the stock charts. Here we further analyze if the opinion-tracking plots for a company affect stock charts, that is, whether the stock price of the next day is affected by the opinions of the current day. However, the experiment shows a precision of only 50.21%.

Figure 9 illustrates the normalized tracking plot and the normalized stock chart of the company BenQ. It shows that opinions have certain short-term relationships with the stock price. However, the time to influence stock price varies. The corresponding points in plots show that this time to influence varies from −1 to 4 days. The time delay to events is because opinions are extracted from the articles published on the Web; the effect of events may persist for several days.

Figure 10 shows the normalized stock chart of the Taiwan Stock Index (a large cap, large-volume indicator of the stock market performance) and Figure 11 shows a normalized randomized stock chart. They are quite different from the tracking plot and the stock chart in Figure 9. Together with

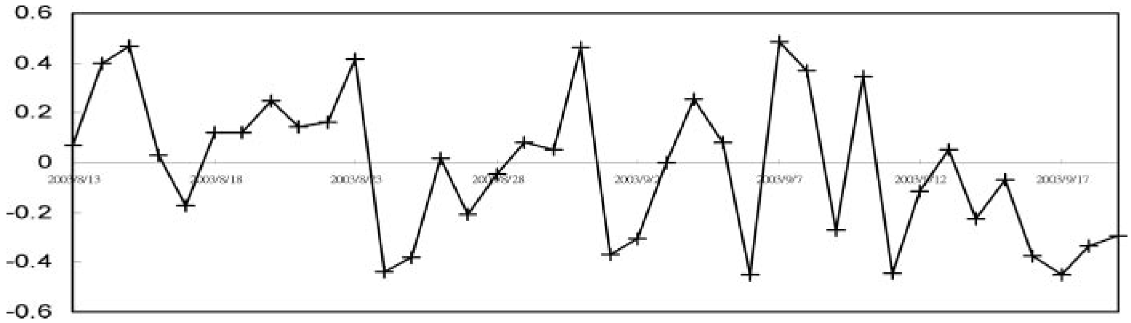


FIG. 11. A random pseudo stock chart.

the satisfactory performance of the curve-overlap related methods, we can infer that opinion tracking plots simulate stock charts. However, opinions are only one of the factors for prediction. Event duration should be considered, and therefore a more complicated model is necessary to predict the time to influence as well as the duration of influence.

Conclusion and Future Work

Our approaches facilitate Chinese opinion mining. Approaches at word, sentence, and document levels function in a bottom-up fashion and can be easily replaced with other approaches to improve system performance in the development phase. Taking into account temporal information when organizing mined opinions allows us to capture and track opinion variations over time.

The way in which we discover relationships between entities leads us to view relationships in terms of related events; our mining of these relationships using opinions about latent events is achievable exactly because of our ability to track opinions. Our experiments extend the objective knowledge base of relationship discovery to the subjective, and we show that even minor changes of opinions—short-term variations of tendencies and weights—are helpful information. This work demonstrates that collocation-based and opinion-based models are complementary, and can be integrated into algorithms that draw from the best of both worlds.

We further gauge the usability of opinions for stock-price prediction, under the assumption that events and opinions influence stock prices. Our analysis shows that opinion variation for a company is related to its stock-price variation: this is encouraging, although we are still far from a good prediction. We have shown that opinions should be taken into consideration when predicting stock prices.

The opinion-analysis system CopeOpi was developed to illustrate applications of the proposed algorithms in opinion mining and relationship discovery. The multicolored and temporal representations of opinions for its IR-system-like user interface provide a clear and friendly visualization for mined opinions and relationships. Moreover, using online news articles as the knowledge base demonstrates the applicability of our proposed algorithms in the real world. Even though the original design of CopeOpi was for mining Chinese opinions, these methodologies can be extended to other languages, which take us to broader research challenges and opportunities in multilingual environments.

Although in this work we discover relationships between companies, our models can be applied to targets other than companies. Homogeneous (company-company, person-person) and heterogeneous (company-person) entities can serve as analysis targets as long as they appear mutually with opinions. However, defining the gold standard and developing evaluation criteria for different types of targets is not trivial.

Systems that support opinion mining have the potential to provide subjective information for end users or further applications, and systems that support relationship discovery have the potential to predict the impact of events on entities. Both of these are indeed worthy goals.

Acknowledgment

This research was supported in part by the National Science Council, Taiwan, under the contract NSC95-2221-E-002-265-MY3.

References

Alm, C.O., Roth, D., & Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language (pp. 579–586). Morristown, NJ: ACL.

Bai, X., Padman, R., & Airoldi, E. (2005). On learning parsimonious models for extracting consumer opinions. In Proceedings of the 38th Annual Hawaii International Conference on System Sciences (Track 3, Vol. 03, p. 75.2). Washington, DC: IEEE.

Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. In Proceedings of the 20th International Joint Conferences on Artificial Intelligence (pp. 2683–2688). Retrieved March 26, 2009, from http://www.cs.cornell.edu/home/cardie/papers/ijcai-2007.pdf

Carenini, G., Ng, R.T., & Pauls, A. (2006). Interactive multimedia summaries of evaluative text. In Proceedings of the 11th International Conference on Intelligent User Interfaces (pp. 124–131). New York: ACM.

Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V.S. (2007). The OASYS 2.0 Opinion Analysis System. Demo in Proceedings of International Conference on Weblogs and Social Media (pp. 313–314). Menlo Park, CA: AAAI.

Chen, K.-H. (2002). Evaluating Chinese text retrieval with multilingual queries. Knowledge Organization, 29(3/4), 156–170.

Chen, K.-H., & Chen, H.-H. (2001). Cross-language Chinese text retrieval in NTCIR workshop—towards cross-language multilingual text retrieval. ACM SIGIR Forum, 35(2), 12–19.

Chinchor, N.A. (1998). Overview of MUC-7. In Proceedings of Message Understanding Conference. Retrieved March 12, 2009, from http://www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html

Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 355–362). Morristown, NJ: ACL.

Dave, K., Lawrence, S., & Pennock, D.M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th International World Wide Web Conference (pp. 519–528). New York: ACM.

Ghose, A., Ipeirotis, P., & Sundararajan, A. (2007). Opinion mining using econometrics: A case study on reputation systems. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (pp. 416–423). Morristown, NJ: Association for Computational Linguistics.

Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. In Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 168–177). New York: ACM.

Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. In Proceedings of the 19th National Conference on Artificial Intelligence (pp. 755–760). Menlo Park, CA: AAAI.

Jindal, N., & Liu, B. (2006). Identifying comparative sentences in text documents. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and development in information retrieval (pp. 244–251). New York: ACM.

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. In Proceedings of the 20th International Conference on Computational Linguistics (pp. 1367–1373). Morristown, NJ: ACL.

Kim, S.-M., & Hovy, E. (2005). Identifying opinion holders for question answering in opinion texts. In Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains. Menlo Park, CA: AAAI.

Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discover and Data Mining (pp. 91–101). New York: ACM.

Ku, L.-W., & Chen, H.-H. (2007). Mining opinions from the Web: Beyond relevance retrieval. Journal of American Society for Information Science and Technology, 58, 1838–1850.

Ku, L.-W., Lee, L.-Y., Wu, T.-H., & Chen, H.-H. (2005). Major topic detection and its application to opinion summarization. In Proceedings of 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 627–628). New York: ACM.

Ku, L.-W., Liang, Y.-T., & Chen, H.-H. (2006). Opinion extraction, summarization, and tracking in news and blog corpora. Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs, AAAI Technical Report (100–107). Menlo Park, CA: AAAI.

Ku, L.-W., Lo, Y.-S., & Chen, H.-H. (2007a). Test collection selection and gold standard generation for a multiply-annotated opinion corpus. In Proceedings of 45th Annual Meeting of Association for Computational Linguistics (pp. 89–92). Morristown, NJ: Association for Computational Linguistics.

Ku, L.-W., Lo, Y.-S., & Chen, H.-H. (2007b). Using polarity scores of words for sentence-level opinion extraction. In Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (pp. 316–322). Retrieved March 12, 2009, from http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/NTCIR/57.pdf

Lin, M.-S., & Chen, H.-H. (2008). Labeling categories and relationships in an evolving social network. In C. MacDonald, I. Ounis, V. Plachouras, I. Ruthven, & R.W. White (Eds.), Lecture Notes in Computer Science, Vol. 4956: Proceedings of the 30th European Conference on Information Retrieval (pp. 77–88). Berlin: Springer.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the Web. In Proceedings of the 14th International World Wide Web Conference (pp. 342–351). New York: ACM.

Liu, Y., Huang, X., An, A., & Yu, X. (2007). ARSA: A sentiment-aware model for predicting sales performance using blogs. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 607–614). New York: ACM.

Manning, C.D., & Schutze, H. (1999). Foundations of statistical natural language processing. Cambridge, MA: MIT Press.

Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C.X. (2007). Topic sentiment mixture: Modeling facets and opinions in Weblogs. In Proceedings of the 16th international conference on World Wide Web (pp. 171–180). New York: ACM.

Mori, J., Ishizuka, M., & Matsuo, Y. (2007). Extracting keyphrases to represent relations in social networks from Web. In Proceedings of IJCAI Conference (pp. 2820–2825). Retrieved March 14, 2009, from http://dli.iiit.ac.in/ijcai/IJCAI-2007/PDF/IJCAI07-453.pdf

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). Mining product reputations on the Web. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discover and Data Mining (pp. 341–349). New York: ACM.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 115–124). Morristown, NJ: ACL.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (pp. 79–86). Morristown, NJ: ACL.

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (pp. 105–112). Morristown, NJ: ACL.

Riloff, E., Wiebe, J., & Wilson, T. (2003). Learning subjective nouns using extraction pattern bootstrapping. In Proceedings of the Seventh Conference on Natural Language Learning (pp. 25–32). Morristown, NJ: ACL.

Robertson, S.E., Walker, S., & Beaulieu, M. (1998). Okapi at TREC-7: Automatic Ad Hoc, Filtering, VLC and Interactive. In NIST Special Publication 500-242: Proceedings of the 7th Text Retrieval Conference (pp. 253–264). Gaithersburg, MD: NIST.

Seki, Y., Eguchi, K., & Kando, N. (2005). Multidocument viewpoint summarization focused on facts, opinion and knowledge. In J.S. Shanahan, Y. Qu, & J. Wiebe (Eds.), Computing Attitude and Affect in Text: Theory and Applications (pp. 317–336). Amsterdam, Netherlands: Springer.

Seki, Y., Evans, D.K., Ku, L.-W., Chen, H.-H., Kando, N., & Lin, C.-Y. (2007). Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access (pp. 265–278). Tokyo, Japan: National Institute of Informatics.

Soboroff, I., & Harman, D. (2003). Overview of the TREC-2003 novelty track. In NIST Special Publication SP 500-255: Proceedings of the 12th Text REtrieval Conference (pp. 38–53). Gaithersburg, MD: NIST.

Stoyanov, V., & Cardie, C. (2006). Toward opinion summarization: Linking the sources. In Proceedings of the Workshop on Sentiment and Subjectivity in Text, ACL-06 (pp. 9–14). Morristown, NJ: ACL.

Takamura, H., Inui, T., & Okumura, M. (2005). Extracting semantic orientations of words using spin model. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (pp. 133–140). Morristown, NJ: ACL.

Wiebe, J. (2000). Learning subjective adjectives from corpora. In Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence (pp. 735–740). Cambridge, MA: AAAI/The MIT Press.

Wiebe, J., Wilson, T., & Bell, M. (2001, July). Identify collocations for recognizing opinions. In Proceedings of the ACL/EACL Workshop on Collocation (pp. 24–31). Retrieved March 26, 2009, from http://www.cs.pitt.edu/∼wiebe/pubs/papers/acl01wkshop.ps

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 347–354). Morristown, NJ: ACL.

Yang K., Yu, N., Valerio, A., Zhang, H., & Ke, W. (2007) Fusion approach to finding opinions in Blogosphere. In Proceedings of the International Conference on Weblogs and Social Media.Retrieved March 14, 2009, from http://www.icwsm.org/papers/2--Yang-Yu-Valerio-Zhang-Ke-new.pdf