Improving Care using Network-Based Modeling and Intelligent Data Mining of Social Media

R. Nalayini
Student
ME Computer Science and Engineering
VPMM Engineering College for Women
Krishnankoil, INDIA
nalayinimurugeshkumar@gmail.com

Mrs. L. Anitha,
M.Sc., M.Tech., (Ph.D).,
Head of the Department
Professor Dept. of Computer Science and Engineering
VPMM Engineering College for Women
Krishnankoil, INDIA
animilk.1982@gmail.com

Mrs. P. Karthika, M.E.,
Assistant Professor
Dept. of Computer Science and Engineering
VPMM Engineering College for Women
Krishnankoil, INDIA
karthikumar.dlp@gmail.com

Mrs. S. Kanaga Lakshmi, M.E.,
Assistant Professor
Dept. of Computer Science and Engineering
VPMM Engineering College for Women
Krishnankoil, INDIA
kanagasvicky@gmail.com

Abstract:- Cleverly extracting information from social media has recently attracted nice interest from the medication and Health science community to at an identical time improve health care outcomes and deflate prices victimization consumer-generated opinion. We've got an inclination to tend to propose a social dancing analysis framework that focuses on positive and negative sentiment, in addition as a result of the aspect effects of treatment, in users' forum posts, and identifies user communities (modules) and influential users for the aim of ascertaining user opinion of cancer treatment. We get a preference to tend to use a self-organizing map to investigate word frequency information derived from users' forum posts. we've got an inclination to tend to then introduced a unique network-based approach for modeling users' forum interactions and utilized a network partitioning technique supported optimizing a stability quality live. This allowed North American nation to work out shopper opinion and establish influential users at intervals the retrieved modules victimization data derived from each word-frequency information and network-based properties. Our approach will expand analysis into showing intelligence mining social media information for shopper opinion of assorted treatments to supply fast, up-to-date data for the pharmaceutical trade, hospitals, and medical employees, on the effectiveness (or ineffectiveness) of future treatments.

Keywords:- *Data mining, social computing , complex networks, semantic web.*

*****

1. Introduction

Social media is adding unlimited opportunities for patients to share their experiences with drugs and devices opportunities, and discuss your business information on their products and services [1] - [3]. Pharmaceutical companies are giving priority to monitoring the social network within their IT departments, creating opportunity rapid dissemination and feedback on products and services to cost optimize and improve performance, turnover and fi pro the increase and decrease [4] . A collection of social media for bio-monitoring data also reported [5]. Social media allows for a virtual network environment. Social media modeling using available network models and computational tools and trends information on fashion documentary 'cloud' of social network structure removed nodes and edges connecting nodes in different relationships. The graphical representation of the most common method of representing information visually can use a series of parties to representations of social network structure to build. Step node can scale the network density and other parameters derived information about the importance of certain entities within the network.

The community groups or modules can Specific algorithms for network group, one of the main tasks in network analysis. Detection requires a specific user communities identify nodes, specific network that enables the extraction of information. Health care providers may use the patient's opinion to improve their services. Doctors may feedback from doctors and other patients get their treatment recommendations and outcomes. Patients may use the information of other consumers in making more informed decisions about their health.

Fig.1. Tree of Rapidminer Processing to ascertain the TF-IDF scores of words in data

Because of the nature of social networks data collection difficult. Various methods have been employed, such as link mining [6] Nomenclature three links [7], predictions based on things [8] Links [9] existence [10] estimates [11] object

[12] Group [13] and detect subgroups [14] and data mining [15], [16].

Predictions, viral marketing, online discussion groups (and ratings) are connected to enable the development of applications based on user feedback solutions. Surveys using traditional social sciences and female subjects in the data collection process, resulting in small sample sizes per study. With social media, content is more readily available, especially when combined with the web and scraping tracking software that enables real-time monitoring of changes within the network. Previous studies used technical solutions to optimize the feeling flu user [17], technology stocks [18], context and sentence structure [19], shopping online [20], classified multiple, government health monitoring, a specific terms on consumer satisfaction, the polarity of the newspaper articles, and user satisfaction evaluation of the companies. Despite the extensive literature, no one has identified the effects users, and how to go forum dynamics of the network relationships. The first phase of our current study, we used the exploratory analysis using self-organizing maps (SOM) to correlations between user and drug messages negative or positive opinion assessment. In the second phase, users and their messages are modeled using a network-based approach. We will build on our previous study and use an improved method for user communities (modules) and impact users displayed in the same manner. This approach seeks current effective potential levels of the organization (scales) within networks and closely watches modules. The approach to reach us Enable the optimal network partition. We subsequently enrich the information retrieval modules from module-frequency word is derived Measures mail users locally and overall user reviews and flag of side effects that may Tarceva, a drug used in the treatment of the raise prevalent cancer: lung cancer.

2. Approaches

2.1. Initial Data Search and Collection

We seek first the most popular cancer boards. We first focused on the number of posts for lung cancer. The table below the number of lung cancer forum jobs: We chose lung cancer because, according to the latest statistics, the most frequently diagnosed cancer in the world for both sexes and the most common cancer in the United States between the sexes. We then compiled a list of drugs that lung cancer patients to see which drugs were most discussed in forums.

Table 1

Forums	Posts on Lung Cancer
cancerforums.net	36523
Cancer-Forums.net	34312
forums.stupidcancer.org	23

It was the drug erlotinib (Tarceva trade name) is the most frequently discussed in forums medicine. Further showed that research Cancerforums.net, despite having slightly fewer posts on lung cancer, erlotinib was dedicated more positions than the other through message boards above. Then we did a search of the drug, using the trade name (Tarceva) and name drug results (erlotinib). The raised more trade name (498) to the name of the drug (66). Search using the trade name back 920 jobs from 2009 to the current date.

2.2. Initial text Mining and Preprocessing

Rapid-Miner scores within each type data collection and processing development trees to search for the positive and negative words common and inverse document frequency of term frequency (TF-IDF). Fig. 1 shows the data collection and processing trees. We origin the first ingredient (Reading Excel) data upload. The downloaded data then the second component ('Data Processing Documents') using subcomponents different (' Extract Content ',' sign A ', respectively,' Counters Filter, 'Stop words filter', 'Cases Transform') variables measurable f processed correctly excessive noise range (misspelled words, words popular stop, etc.), screening available.

$$w_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d} & \text{if } tf_{t,d} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Final component ('processed data') with each word containing TF-IDF specific score, run the final word list. TFT, otherwise equal to the case ≥1 TFI in other words if the weight, IT d 1) log n XT 0: Then weight allocated to each of the words in a user message document using the following formula (d) frequency (t) represents the number of documents in the entire collection and a warm place XT document number.

2.3. Cataloging and Tagging Text Data

A tagged text data which scores highest TF-IDF personnel NLTK vehicles (http://www.nltk.org/) MATLAB they use negative words to make sure that is reflected is negative and context positivity positive words. This approach was used before using negative tags on positive word.

Table 2
Wordlist for Post-Analysis

Positive	Negative
Good	Fear
Enjoy	Bad
Favourable	Don't
Grateful	Hate
Feasible	Hurt
Effective	Fail
Great	Painful
Favourably	Difficult
Help	Disagree

We received a positive label of negative words. We NLTK analysis tools used and the application of water, in the true sense of the word match contextual settings. For example, the context of the term before that (in our case 'great is

labeled as a negative as there is sufficient labeled 'must' so 'I do not feel good' as should have to include statements that) has returned to locate the specific f. Das and Chen classify words [18] used a similar approach. We went a step further and think positive label of negative words. No side effects so I'm happy, "he's an understatement! before returning it to the fi specifications state the word "as a result of 'not' (positive context).

Table 3
List of Side Effect Words

Rash
Weakness
Nausea
Headache
Itching
Acne
Vomiting

We preface method final word list using pruned some of the words positive and negative, statistically insignificant multiplier measurements to obtain a uniform set and finish.

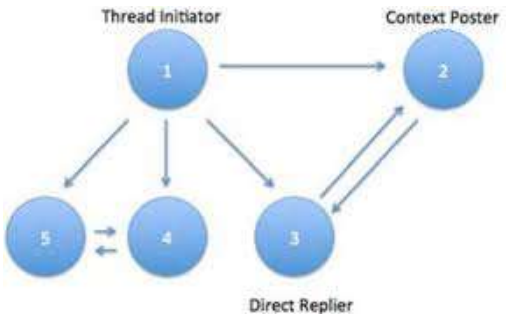

Fig 2. Thread model where nodes represent users/posts and the edges represent information transferred among users.

We seem to us less than ten times each word permission was expelled parallel procedures final results are shown in Table II of the list of words with some editing changes (55 positive and 55 negative words words) 110 words automatically to look for side effects of erlotinib move the user's work. goal, we controlled the word Medicine, National Library of Medical Subject Title used.

Then in hand (online dictionaries, such as Merriam-Webster using Synonyms control reduced the number of words (http://www.merriam-webster.com/) and automatically (database software database thesaurus as a synonym (http://www.language-databases.com/) and the results are shown in Table fi with NDR synonym search Google.

2.4. Consumer Sentiment Using a SOM

Analysis using this part of SOMA's Consumer Confidence, All posts tagged exploratory analysis via SOMs. The manual allowed us written manual verification method results can be used as positive and negative before feeding the data collected by the general opinion of the cause of. High-dimensional data SOMs are neural networks that have low-dimensional representation. Within the network, a layer represents output space which is assigned a specific weight. Weight values to reflect the cluster material. SOMs together have similar weight data such as neurons displays, network data.

2.5. Consumer Sentiment Using a SOM

Some Consumer Confidence analysis using this section, all posts are due to data collected by public opinion before feeding manual verification method results were used as positive and negative labels written exploratory analysis via SOMs. The manual labeling allowed us for High-dimensional data SOMs are neural networks that have a low-dimensional representation. In the network, a layer represents output space with each neuron assigned a specific weight. Such neurons displays network data as SOMs similar weight data.

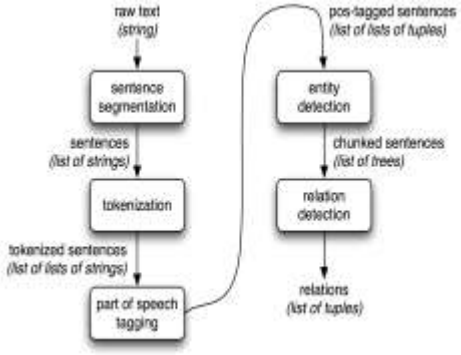

Fig 3. Figure illustration of the framework of our network-based analysis. Initially, the posts retrieved from the forum via Rapid-miner are preprocessed using the NTLK Toolbox.

2.6. Modeling Forum Postings using Network Analysis

Discovering influential users was future step in our analysis. to the current goal, we tend to designed networks from forum posts and their replies, whereas accounting for content-based grouping of posts ensuing from the prevailing forum threads. Networks square measure composed of nodes and their associations: they're either non-directional (a connection between 2 nodes while not a direction) or directional (a reference to AN origin and an end).

$$Q_{M_t} = \frac{1}{2m}\sum_{i,j}\left(A_{t_{ij}} - \frac{d_i d_j}{2m}\right)\cdot\delta(i,j),$$

The nodal degree of the latter measures the quantity of connections from the origin to the destination. Four node sorts are identified among a network: Isolated, transmitter, receptor, and carrier. The network's density measures this range of connections.

The network-based analysis is wide utilized in social network analysis supported its ability to each model and analyze inter-social dynamics. We tend to devise a directional network model owing to the character of the forum beneath scrutiny (multiple threads with multiple thread initiators) and its internal dynamics among the members (members reply to string initiators further on alternative users).

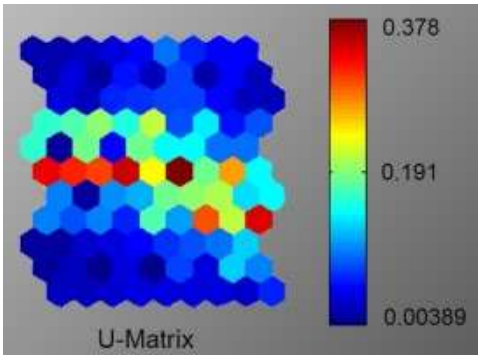

Fig. 4. The diagram of U-Matrix of the posts from the Cancerforums.net forum.

A pair of describes the approach we tend to selected to create our network that shows however every posting-reply combine is modeled. Fig.4 shows the U-Matrix of the posts from the Cancerforums.net forum.

2.7. Identifying Sub-graphs

Our modeling framework has consequently born-again the forum posts into many massive directional networks containing variety of densely connected units (or modules) (see Fig. 3, step A1). These modules have the characteristic that they're a lot of densely connected internally (within the unit) than outwardly (outside the unit). We have a tendency to select a multi scale technique that uses native and world criteria for distinguishing the modules, whereas maximizing a partition quality live known as stability. The soundness live considers the network as a Markov process, with nodes representing states and edges being attainable transitions among these states. In , the authors projected Associate in nursing approach within which transition chances for a stochastic process of length t (t being the mathematician time) modify multi scale analysis. With increasing scale t, larger and bigger modules square measure found.

At is that the contiguousness matrix, t is that the length of the network, m is the number of edges, i and j are nodes, di is node i's (and j's) strength, and δ (i,j) perform becomes one if one in all the nodes belong to constant network and 0 if it doesn't belong to any network. At is computed as follows (in order to account for the random walk): At = D•Mt, wherever M = D−1 •A (D being the square matrix containing the degree vector giving for every node its degree) [28]. The strategy for distinguishing the best modules is predicated on alternating native and world criteria that expand modules by adding neighbor nodes, reassigning nodes to completely different modules, and significantly overlapping modules until no further optimization is possible, in line with (1). Many partitioning schemes were obtained unfinished on the range of scales employed by the method, with the optimal partitioning having the biggest stability. We have a tendency to name the modules therefore retrieved data modules

2.8. Module Average Opinion and User Average Opinion

We then precede to refine the knowledge modules through feeding them with the knowledge obtained from the forum posts (using the wordlist vectors). In a very first step, we have a tendency to gear toward distinctive influential users at intervals our networks. Influential users square measure users that broker most of the knowledge transfer at intervals network modules and whose opinion in terms of positive or negative sentiment towards the treatment is 'spread' to the opposite users at intervals their containing modules. to the current goal, we have a tendency to enriched the knowledge modules obtained as delineate in Section II-F with the TF-IDF a lot of the user posts resembling the users found in every module. The TF-IDF scores from the wordlist of positive and negative words (see Table I) were wont to build 2 sorts of measure. The world live (pertaining to the full data module) is depicted by the module average opinion (MAO). It examined the TF-IDF a lot of postings matching the nodes in a very specific module

MAO = Sum+ −Sum− Sum all

Sum+ = xij is that the total total of the TF-IDF scores matching the positive words within the wordlist vectors at intervals the module. The units i represent post index. The unit j represents the wordlist index (matching the positive words within the list). Sum− = Xij is that the total of the TF-IDF scores matching the negative words within the wordlist vectors at intervals the module. The units i represent post index. The unit j represents the wordlist index (matching the negative words within the list). Total all = N i=1 M k=1 xik is that the total of each of the fore mentioned sums. The unit k is that the index running across variables throughout the complete wordlist. The native live that illustrates specific user opinion to every node within the module (the user average opinion, or UAO) that examines the TF-IDF scores to the common of the collected posts of the user is that the following:

UAOi = Sumi+ Sumi / Sumiall

Sumi+ = j∈P xij is that the TF-IDF score's total matching to positive words for the ith user's wordlist vector. P is that the index set denoting the wordlist's positive variables.

Sumi− = j∈N xij is that the TF-IDF score's total matching to negative words for the ith user's wordlist vector. N is that the index set denoting the wordlist's negative variables. Total all = M j=1 Xij is that the total of each sums, and j is that the index of the full wordlist. H. data Brokers at intervals the knowledge Modules we have a tendency to first hierarchic individual nodes in terms of the overall variety of connecting edges (in and out-degree) to spot influential users at intervals the modules. We have a tendency to then looked nodes in every module supported the subsequent criteria: 1) the nodes have densest degrees at intervals the module (highest variety of edges). 2) The UAO scores equate the signs of the MAO of the containing module. The nodes that qualified were dubbed data brokers, supported the said criteria. Their giant nodal degrees guarantee increased data transfer compared to alternative nodes whereas their matching UAO and MAO scores reflect consistency of positive or negative opinion at intervals the containing module.

2.9. Network-Based Identification of Side Effects

In the second step of our network-based analysis, we have a tendency to devise a method for distinguishing potential aspect effects occurring throughout the treatment and that user posts on the forum highlight. to the present goal, we have a tendency to overlay the TF-IDF innumerable the second wordlist (see Table II) onto modules obtained in Section II-F. The TFIDF scores inside every module can so directly reflect however frequent an explicit side-effect is mentioned in module posts. afterward, a applied mathematics take a look at (such because the t-test for example) may be accustomed compare the values of the TF-IDF scores inside the module to those of the forum population and determine variables (side-effects) that have significantly higher scores. Fig. three presents a diagram that visually describes the steps in our network-based analysis.

3. Results

A set consisting of half-hour of the info was used for coaching the monetary unit. we have a tendency to use a 12×12 map size with one hundred ten variables comparable to the positive and negative terms to determine the burden of the words corresponded to the opinion of the drug Erlotinib. As mentioned in Section II, every word from the list appeared over 10 times. This achieved a homogeneous measuring set whereas eliminating statistically insignificant outliers. Abundant of the user's posts converged on 3 of the map. we have a tendency to checked the several nodes' correlation with their weight vectors' values comparable to positive or negative words to define the positive and negative areas of the map. The user opinion of Erlotinib was overall satisfactory, with Table III summarizing the satisfaction/dissatisfaction below: in step with chart, and from our readings of each the user posts and also the Kyrgyzstani monetary unit, the foremost pressing concern from each camps was the facet effects, that area unit extensively documented within the medical literature. The prices of the drug were conjointly another matter of concern (albeit limited). We have a tendency to then precede to spot influential users. Our modeling approach yielded at first one loosely connected network, linking all users at intervals the forum. subsequent module identification mistreatment the strategies delineate in Section II-F yielded associate best partitioning containing five densely connected module. We have a tendency to varied our scale parameter at intervals the interval t [0,2] in 0.1 increments, as prompt by. Varied the dimensions parameter resulted in a very set of partitions starting from modules supported single individual users (for scale parameter t =0), to massive modules (for values of t near the higher limit of the interval).

Table III

Opinion from users of ERLOTINIB

Satisfaction	Dissatisfaction
80 percent	20 percent

Breakdown of User Opinion

Fully Satisfied (31)	Full Dissatisfaction (6)
Satisfied Despite Side Effects (26)	Dissatisfaction because of Side Effects (18)
Satisfied Despite Costs (12)	Dissatisfaction because of Costs (9)

The best partition (maximizing the standard live in (1) was obtained for t =1. On the Cancerforums.net message board, 10 users out of the 920 posts were identified as info brokers as shown in Fig. 5(a)–(e) below. Densities of the retrieved modules vary from zero.2 to 0.6.

These density values were at intervals the ascertained density values interval (towards the higher limit), when put next to those usually noted in social networks, so confirming our network modeling approach. Information brokers were identified following the procedure delineate in Sections II-G–H. any scrutinizing these users and their containing modules we have a tendency to confirmed their connections were the densest. a radical reading of those 10 users' posts throughout the threads they started and took part in disclosed that they were informative and actively interacting with users across several threads.

Different members wanted out these 10 posters for his or her knowledge and skill. Their forum 'behavior' has confirmed to North American country that these users were the premier info brokers of the medicine Erlotinib on the Cancerforums.net forum. In the last a part of our analysis, we have a tendency to investigate that modules were significantly concerned in discussing specific facet effects.
As delineate in Section II-I, retrieved modules were enriched with the TF-IDF scores comparable to the side-effect word list vectors. For every module and every side-effect scores sample, tests were performed to assess the significant distinction between the in-module sample and also the overall forum population scores. Rash and haptic sensation were identified because the facet result terms with significantly higher scores in Modules one, 2, and five when put next to the scores population within the forum.

This reflects the very fact that users sorted at intervals these modules repeatedly mentioned these facet effects in their posts. This was confirmed by subsequent scrutiny of the several posts. A literature search confirmed that rash and haptic sensation area unit so 2 of the foremost common side-effects of Erlotinib with as much as 70% of the patients affected, as indicated by clinical studies.

4. Discussion

We have a tendency to regenerate a forum centered on medicine into weighted vectors to live client thoughts on the drug Erlotinib mistreatment positive and negative terms aboard another list containing the facet effects. Our strategies were able to investigate positive and negative sentiment on carcinoma treatment mistreatment the drug by mapping the massive dimensional information onto a lower dimensional area mistreatment the monetary unit. Most of the user information was clustered to the world of the map connected to positive sentiment, so reflecting the final positive read of the users. Subsequent network primarily based modeling of the forum yielded fascinating insights on the underlying info exchange among users. Modules of

powerfully interacting users were identified employing a multi scale community detection methodology delineate.
By overlaying these modules with content-based info within the variety of word-frequency scores retrieved from user posts, we have a tendency to were able to determine info brokers that appear to play vital roles within the shaping the data content of the forum. to boot, we have a tendency to were able to determine potential facet effects systematically mentioned by teams of users. Such associate approach may well be wont to raise red flags in future clinical police investigation operations; likewise as lightness numerous different treatments connected problems.

The results have opened new prospects into developing advanced solutions, likewise as revealing challenges in developing such solutions. The accord on Erlotinib depends on individual patient expertise. Social media, by its nature, can bring totally different people with different experiences and viewpoints. We have a tendency to sift through the info to find positive and negative sentiment, that was later confirmed by analysis that emerged relating to Erlotinib's effectiveness and facet effects. Future studies would require additional u p-to-date info for a clearer image of user feedback on medicine and services. Future solutions would require additional advanced detection of repose social dynamics and its effects on the members: such interests of study might embrace rankings, 'likes' of posts, and friendships. Any stress on context posting would require formal language dictionaries that embrace medical terms for specific diseases, and informal language terms ('slang') to clarify posts.

Finally, completely different platforms can permit up-to-date info on the standing of the drug just in case one social platform ceases to debate the drug. Another answer will examine multiple word lists that may embrace multiple treatments that, once combined with discourse posting and medical lexical dictionaries, will pinpoint the supply (or multiple sources) of user satisfaction (or dissatisfaction), which might open the door towards mapping client sentiment of multi drug therapies for advanced diseases. This answer are often visualized on future medical devices that function post selling feedback circuit that customers can use to precise their satisfaction (or dissatisfaction) on to the corporate.

The corporate benefits from period of time feedback that may then be wont to assess if there are a unit any issues and speedily address such issues. Social media will open the door for the health care sector in address price reduction, product and repair improvement, and patient care.

5. References

[1] A. Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. "Artificial societies and social simulation using ant colony, particle swarm optimization and cultural algorithms," New Achievements in Evolutionary Computation, P. Korosec, Ed. Rijeka, Croatia: InTech, pp. 267–297, 2010.
[2] W. Cornell and W. Cornell. (2013). How Data Mining Drives Pharma: Information as a Raw Material and Product. [Online]. Available: http://acswebinars.org/big-data
[3] L. Toldo, "Text mining fundamentals for business analytics," presented at the 11th Annual Text and Social Analytics Summit, Boston, MA, USA, 2013.
[4] L. Dunbrack, "Pharma 2.0 – social media and pharmaceutical sales and marketing," in Proc. Health Ind. Insights, 2010, p. 7.
[5] C. Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural data mining of influenza mentions in web and social media," Int. J. Environ. Res. Public Health, vol. 7, pp. 596–615, Feb. 2010.
[6] L. Getoorand C. Diehl," Link mining: a survey,"SIGKDDExplor. Newsl., vol. 7, pp. 3–12, Dec. 2005.
[7] Q. Lu. And and L. Getoor, "Link-based classification," in Proc. 20th Int. Conf. Mach. Learning, Washington, D.C., USA, 2003, pp. 496–503.
[8] A. Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in Proc. SIGIR Conf. Inform. Retrieval., New Orleans, LouisianaLO, USA, 2001, pp. 258–266.
[9] B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," in Proc. Adv. Neural Inform. Process. Syst., Vancouver, B.C. Canada, 2003.
[10] D. Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," J. Am. Soc. Inform. Sci. Technol., vol. 57, pp. 556–559, May 2007.
[11] Z. Lacroix, H. Murthy, F. Naumann, and L. Raschid, "Links and paths through life sciences data sources,"inProc.1stInt.WorkshopDataIntegr. Life Sci., Leipzig, Germany, 2004, pp. 203–211.
[12] J.Noessner, M.Niepert, C.Meilicke, and H.Stuckenschmidt, "Leveraging terminological structure for object reconciliation," in The Semantic Web: Research and Applications, Heidelberg, Berlin: Springer, 2010, pp. 334– 348.
[13] M. E. J. Newman, "Detecting community structure in networks," Eur. Phys. J., vol. 38, pp. 321–330, Mar. 2004.
[14] J. Huan and J. Prins, "Efficient mining of frequent subgraphs in the presence of isomorphism," in Proc. 3rd IEEE Int. Conf. Data Mining, Melbourne, Florida.FL, USA, 2003, pp. 549–552.
[15] D. Hand, "Principles of data mining," Drug Safety, vol. 30, pp. 621–622, Jul. 2007.
[16] J. Hans and M. Kamber, Data Mining: Concepts and Techniques. 2nd ed. Burlington, MassMA, USA: Morgan Kaufmann, 2006
[17] C. Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural data mining of influenza mentions in web and social media," Int. J. Environ. Res. Public Health, vol. 7, pp. 596–615, Feb. 2010.
[18] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the Web," Manag. Sci., vol. 53, pp. 1375–1388, Sep. 2007.
[19] E. Riloff, "Little words can make a big difference for text classification," in Proc.18th Annu. Int. ACMSIGIR Conf. Res. Develop. Inform.Retrieval, Seattle, Washington WA, USA, 1995, pp. 130–136.
[20] W. Yih, P. H. Chang, and W. Kim, "Mining online deal forums for hot deals," in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., Beijing, China, 2004, pp. 384–390.