A Text Cube Approach to Human, Social and Cultural Behavior in the Twitter Stream

Xiong Liu[1], Kaizhi Tang[1], Jeffrey Hancock[2], Jiawei Han[3], Mitchell Song[1], Roger Xu[1], and Bob Pokorny[1]

[1] Intelligent Automation, Inc.
[2] Cornell University
[3] University of Illinois at Urbana-Champaign
xliu09@gmail.com, {ktang,msong,hgxu,bpokorny}@i-a-i.com,
jeff.hancock@cornell.edu, hanj@illinois.edu

Abstract. Twitter is a microblogging website that has been useful as a source for human social behavioral analysis, such as political sentiment analysis, user influence, and spread of news. In this paper, we discuss a text cube approach to studying different kinds of human, social and cultural behavior (HSCB) embedded in the Twitter stream. Text cube is a new way to organize data (e.g., Twitter text) in multiple dimensions and multiple hierarchies for efficient information query and visualization. With the HSCB measures defined in a cube, users are able to view statistical reports and perform online analytical processing. Along with viewing and analyzing Twitter text using cubes and charts, we have also added the capability to display the contents of the cube on a heat map. The degree of opacity is directly proportional to the value of the behavioral, social or cultural measure. This kind of map allows the analyst to focus attention on hotspots of concern in a region of interest. In addition, the text cube architecture supports the development of data mining models using the data taken from cubes. We provide several case studies to illustrate the text cube approach, including public sentiment in a U.S. city and political sentiment in the Arab Spring.

1 Introduction

Human Social Cultural Behavior (HSCB) analysis and modeling is an emerging research area that focuses on understanding, predicting, and shaping human behaviors cross-culturally [1]. As social media becomes more prevalent, HSCB analysis can take advantage of the availability of in-situ data for real-world applications. Nowadays, many social media sites provide Application Programming Interface (API) invocation via web services. For example, Twitter is a microblogging website that has been useful as a source for HSCB analysis (e.g., political sentiment analysis [2], user influence [3], and spread of news [4]). The Twitter streaming API allows applications to have real-time access to tweet objects. Using this API, we can design code to automatically extract live tweets for a topic, transform and load them into the textual database for subsequent analysis.

A.M. Greenberg, W.G. Kennedy, and N.D. Bos (Eds.): SBP 2013, LNCS 7812, pp. 321–330, 2013.

In this paper, we introduce a text cube approach to social media analysis, especially sentiment analysis. In data warehousing, data cube is a way to organize data in multiple dimensions and multiple hierarchies for information query and visualization from multiple perspectives [5]. A data cube allows data to be aggregated and viewed from multiple perspectives, and it is defined by measures and dimensions. The measures (or facts) are numeric values that are usually additive (e.g., sales of a product). Analysts need to look at measures using some "by" conditions. The "by" conditions are dimensions. For example, in order to analyze sales volume, analysts often want to see its measure by day and by location. In this sense, dimensions are the perspectives with respect to which an analyst wants to aggregate or view measures. The advantage of using the data cube approach over using the relational database approach to organize multidimensional data is that data cube performs very well with complex queries and analysis of very large datasets [5].

A data cube can be extended to summarize and navigate structured data together with unstructured text. Such a cube is called text cube [6,7]. Unlike a traditional data cube where measures are directly retrieved from the original databases, text cube provides an advanced text analytics capability for extracting HSCB measures from unstructured text streams. With the HSCB measures defined we are now able to view the text cube and perform analyses. This includes slicing, dicing and drilling through cube cells. Text cube also supports complex analysis methods, such as topic modeling [8], online analysis [9], and keyword-based exploration [10].

We have added two additional functionalities beyond the standard text cube methodology. First, we have added the capability to display the contents of the cube on a heat map. Basically, the heat map shows each geographic region with a shade of red. The degree of opacity is directly proportional to the value of the measure; the larger the measure is the more opaque the color. This kind of map allows the analyst to focus attention on hotspots of concern in the region of interest. Second, we have added support for the development of prediction models using a data mining approach [11,12]. Therefore, text cube is a practically useful approach for organizing and analyzing text-based communications to assess HSCB dimensions (e.g., sentiment or affect, deception, group identity) for a given group and to predict current belief states and likely intended actions.

This paper is organized as follows. We first describe the text cube approach to HSCB analysis. Then we present several case studies to illustrate the text cube approach, including public sentiment in a given region and political sentiment in the Egyptian Revolt. Finally, we conclude the paper and discuss future research.

2 A Text Cube Approach

We have developed a dynamic data cubing and mining framework, called *SocialCube* [6], for large amounts of HSCB data. SocialCube is an advanced data cube architecture that allows analysts to summarize and navigate structured data together with unstructured text for efficient query and analysis. In SocialCube, linguistic feature analysis is a preliminary step for developing text-based data cubes or text cubes.

2.1 HSCB Linguistic Analysis for Sentiment

We have designed a comprehensive HSCB linguistic feature analysis framework that allows for an extensible set of HSCB dimensions, such as affect/sentiment [13], deception [14], sense of fatalism vs. mastery, and power structure. Here we focus on the linguistic feature analysis of sentiment.

Perhaps one of the most important social and cultural dynamics for humans is their sense of emotion [15]. Emotion reflects not only how an individual is reacting to ongoing events, but can also reflect to how an individual generally views the world and his/her place in it. While emotion was long ignored by cognitive psychologists, a wide preponderance of data suggests that understanding an individual or group's emotional state can provide important insight and prediction into their decision-making, cognitive responses, and future behavior [16].

Although emotion is often assumed to be only communicated nonverbally [17], a number of recent studies suggest that humans convey their emotions in text-based communication, such as emails, blogs, instant messaging, and other forms of textual communication through linguistic cues. In one study [18], for example, individuals were asked to communicate only by text, and one partner was induced to feel sad before the interaction. Under these conditions, their partner was able to detect the negative emotion in the emotionally induced participant, indicating that emotion can be detected in text-based communication. Importantly for the present research, these data suggest that emotions can be detected from text-based communication.

There are specific linguistic patterns of emotional expression in verbal content, and there are a number of established tools that can extract relevant emotional content, including the Linguistic Inquiry and Word Count program and the Dictionary of Affect in Language program. Using these tools, Hancock and colleagues [19] have found that when people are sad they tend to use fewer words, disagree more, use more negative-affect words, and respond more slowly.

These kinds of verbal patterns are extractable not only at the dyadic or group level, but also at the organization and even national level. Consider Kramer's work [20] on Facebook status updates and his assessment of the Gross National Happiness index. The Gross National Happiness index assesses the emotional context of the United States by extracting positive and negative emotional indicators from 100 million Facebook users. This analysis, based on the textual content from status updates, correlates highly with self-reported satisfaction as well as culturally and emotionally significant calendar events (e.g., Christmas, death of a politician, etc.). A more recent study provides even more powerful evidence that affect words from tweets reflect actual emotional states. Golder and Macy [21] analyzed positive and negative affect words from 550 million tweets collected from cultures around the world. Their data revealed that both positive and negative affect in tweets tracked precisely with daily circadian rhythms, with positive affect peaking in the mid-morning and mid-evening, and negative affect peaking mid-afternoon and very early morning.

Taken together, these data suggest that emotional indicators or an individual or group can be extracted from verbal content present in text-based communication, and that these features, dynamically tracked over time, can predict emotionality of an individual or even a group.

In our current research, we use the Linguistic Inquiry and Word Count (LIWC) tool [22] to extract emotional features. LIWC counts word frequency along the approximately 65 dimensions of language in the default LIWC dictionary. These dimensions include function word categories, such as pronouns, articles and auxiliary verbs, as well as psychological categories, such as affect and cognition-related words, and social dimensions, such as family words. The output for each word category from the LIWC default dictionary represents a feature in our analysis. Note that some LIWC measures have hierarchical relationships. For example, "affective processes" can be divided to "positive emotion" and "negative emotion"; and "negative emotion" can be further divided into "anger", "anxiety", and "sadness".

2.2 Text Cube Construction for Sentiment

Our goal is to organize linguistic features/indicators of sentiment in a data cube model, a new way to organize data in multiple dimensions and multiple hierarchies for efficient information query and visualization from multiple perspectives [7]. A data cube allows data to be aggregated and viewed in multiple dimensions. It is defined by dimensions and facts (or measures). In general terms, dimensions are the perspectives with respect to which an organization wants to keep records (e.g., by time, by location, etc.). Each dimension may have a table associated with it called a dimension table. Facts are numerical measures that are quantities by which we want to analyze relationships between dimensions.

The star schema is a multidimensional data model to design the data cube. In a star schema, there are one or more fact tables referencing any number of dimension tables. The fact table contains the names of the facts (measures), as well as keys to each of the related dimension tables. We have designed a star schema to store the extracted linguistic features for different HSCB dimensions including sentiment. They are stored as measures in the Fact table. Based on star schema, we have designed a data cube architecture to allow users to conveniently view aggregated statistics of sentiment relevant measures along different dimensions, such as time and location.

3 Case Studies on Cubing and Visualization

Several data cubes have been designed and implemented for demonstration purposes, including one cube looking at tweets originating from the Washington, D.C. region, and one cube looking at tweets on the topic of the Egypt revolt.

3.1 Affective Processes Cube in Washington, D.C.

We collected ~0.5 million tweets in the Washington D.C. region for the period of May to July 2011. **Fig. 1** shows the database screen shot with sample tweets. Then we performed HSCB linguistic analysis of the tweets to extract sentiment measures, such as "positive emotion" and "negative emotion". The extracted measures and other structured information are stored in a star schema. The dimensions for this schema are Date, Zone, User, Location, Event, and Tag, and are represented by database tables date_dim, zone_dim, user_dim, location_dim, event, and tags, respectively. The facts (i.e. measures) are stored in the fact table pycholinguistic_facts.

date	time	tweet_text
2011-05-21	17:22:45	My Mom said, "if anything weird happens for the 'end of the world', come to the house..." "uhhh...oh...kayyyy"...
2011-05-21	11:46:38	An Igbo man fell into a well one day & started screaming for help. His wife rushed off to buy a rope to sav... (cont)...
2011-05-21	19:12:50	@NickyyBarness naw a clothin joint
2011-05-21	08:52:45	@OzzieNeutron awwwe that's nice
2011-05-21	18:11:01	Did Medieval Times for dinner. Was a great show and big fun.
2011-05-21	14:12:15	I'm at Union Street Public House (121 S Union St, at Prince St, Alexandria) http://4sq.com/iTNObs
2011-05-21	11:00:27	#endoftheworldconfessions i keep a roll of toilet paper in my truck in case i have to go in a public bathroom and ...
2011-05-21	09:11:18	So who went to bed last nite thinking the world was gonna end today?
2011-05-21	08:43:32	I have heard good things about Xoom too @RdLessTkn Until I use them I am not the best to decide
2011-05-21	15:04:53	BRAVA Heyyy! @ JussCallMeOzz
2011-05-21	13:07:12	I'm at Fiesta Asia Street Fair (Pennsylvania Avenue NW, btw 3rd & 6th, Washington) w/ 6 others http://4sq.com...
2011-05-21	10:22:08	When I get my abs I'm going running shirtless.
2011-05-21	12:09:02	Chilling in da crib @ heathaplexVISION http://gowal.la/c/4gHzZ
2011-05-21	15:57:05	We get it by now

Fig. 1. Sample tweets

In order to view and perform analysis on the data, a cube must be defined and the dimensions and measures mapped to corresponding database table columns. This is done in an XML file referred to as a cube schema. A cube schema contains a logical model, consisting of cubes, hierarchies, and members, and a mapping of this model onto a physical model. With the schema defined we are now able to view the tweets cube and perform analysis using an Online Analytical Processing (OLAP) tool. This includes slicing, dicing and drilling through cells.

Fig. 2 is a cube interface of the Affective Processes Cube for the Washington, D.C. Region with Date as the horizontal dimension and Zone as the vertical dimension. Zone, as we mentioned, represents a rectangular area on the Earth. Each measure is associated with a zone as determined by the tweet location (i.e. latitude and longitude the tweet originated from). Measures are shown along the horizontal axis. **Fig. 2** shows the average negative emotion for each day for each zone. The cube interface can display a different view of the same cube by showing the average measure for all days combined for each of the zones. The cube interface also supports many charting options (e.g., 3D vertical bar chart).

	Date					
	2011-05-21	2011-05-22	2011-05-23	2011-05-24	2011-05-25	2011-05-26
	Measures	Measures	Measures	Measures	Measures	Measures
Zone	Negative Emotion	Negative Emotion	Negative Emotion	Negative Emotion	Negative Emotion	Negative Emotion
All Zones	2.469	1.526	0.985	1.197	1.079	1.233
unknown	3.05	1.78	0.9	1.161	1.091	1.322
Zone_1	2.684	1.486	0.819	1.289	1.667	2.145
Zone_1_3	0	5		0		0
Zone_1_4	2.765	1.216	0.819	1.428	1.667	2.264
Zone_2	1.308	1.316	1.714	0	0.94	0.275
Zone_2_2		0				
Zone_2_3	1.379	1.513	1.818	0	1.019	0.275
Zone_2_4	0	0	0	0	0	
Zone_3	3.72	0	0.59	0.909	0.312	2.077
Zone_3_1					12.5	
Zone_3_2	3.469	0	0.616	1.186	0	2.439
Zone_3_3						
Zone_3_4	5.557	0	0	0	0	0
Zone_4	2.036	1.374	1.038	1.314	1.124	1.113
Zone_4_1	2.098	1.363	1.065	1.329	1.154	1.093
Zone_4_2	0.392	1.838	0	1	0	0.769
Zone_4_3	4.763			4.76		
Zone_4_4	0		0	0	1.111	1.704

Fig. 2. Affective Cube for individual dates

3.2 Heat Map Visualization

Along with viewing and analyzing tweets data using cubes and charts we have also added the capability to display the contents of the cube on a heat map. **Fig. 3** is a representative heat map. It shows each zone with a shade of red. The degree of opacity is directly proportional to the value of the measure; the larger the measure is the more opaque the color. The blue area on the right lists the map type options as well as the overlays that may be turned on or off by either checking or unchecking the respective boxes. Each overlay represents a column of the cube currently being displayed. For example, the first overlay (after the Show Grid overlay) is the first column from **Fig. 2** for the May 21, 2011 date.

If a measure is in the lower 25 percent range, its corresponding zone is given an opacity level of 0.2 (light pink). If a measure is within the 25 percent and 50 percent range, its corresponding zone is given an opacity level of 0.4. If a measure within the 50 percent and 75 percent range, its corresponding zone is given an opacity level of 0.6. Finally, zones of measures above the 75 percent are given an opacity level of 0.8 (red). In this example, the heat map allows the analyst to quickly visualize the areas that have the highest negative emotions on a given day or set of days.

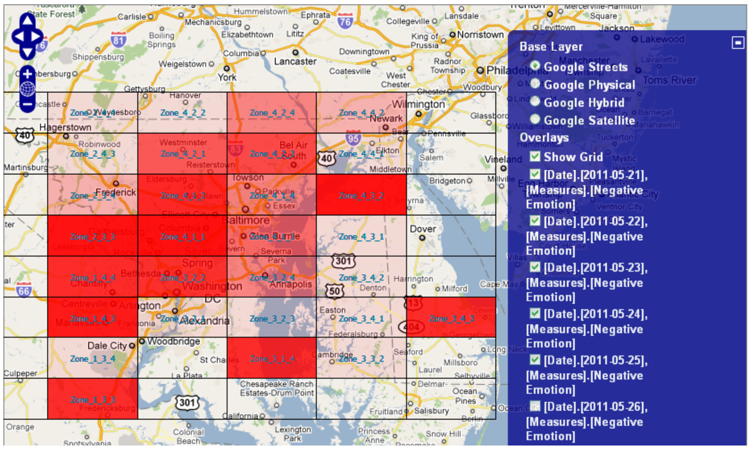

Fig. 3. Heat map interface

In addition, we have studied the affect (e.g., negative emotion) change in a given region. **Fig. 4** shows the heat map plot of negative emotion in the Washington DC area. We note that Zone 4-3-2 has relative high negative emotion on both May 21 and May 24, suggesting that these regions may be of interest for further investigation, such as correlating the emotion with known events in the area, crime statistics, or even weather conditions. Schwarz and Clore (1983) [23] showed that peoples' emotional state is driven by the weather in their location, but that emotion reports would return to "normal" once the weather's influence was pointed out. In other words, the weather may be an unidentified cause of emotion at the aggregate level that to date has not been fully investigated.

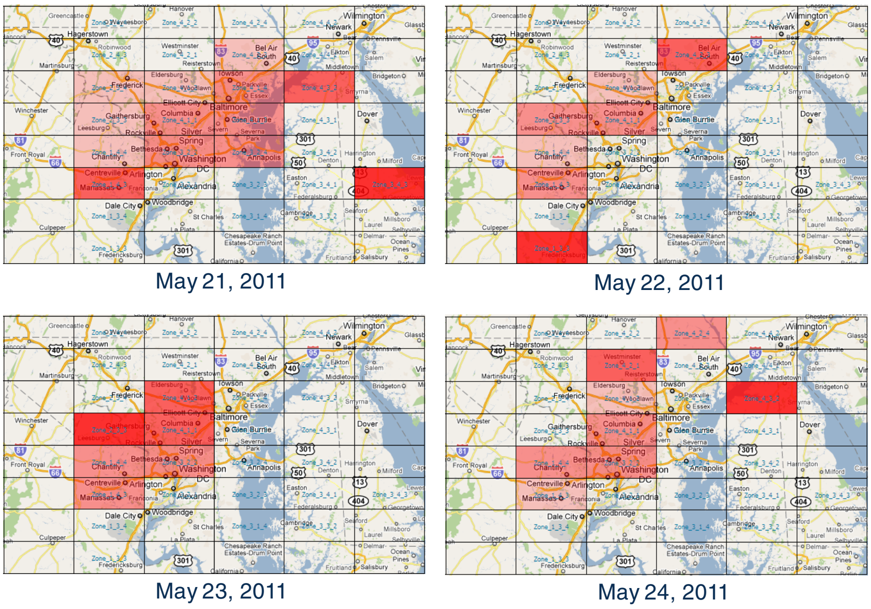

Fig. 4. An example of negative emotion change

3.3 Affective Process Cube for the Egypt Revolt

As an example of political sentiment, we built an Affective Processes cube for the Egypt revolt (see **Fig. 5**). The horizontal dimension is the cities from which the tweets originated from and the vertical dimension is the time period. Its measures include Affective Processes, Positive Processes, Negative Processes, Anger, Anxiety, Sadness, Religion, and Social. We can expand the time dimension to drill down to specific days. We can also aggregate the sentiment measures for all locations by "shrinking" the location dimension. This will generate the dataset for trend analysis of sentiment regardless of locations.

	Time							
	+All Periods							
	Measures							
Location	Affective Processes	Positive Emotion	Negative Emotion	Anger	Anxiety	Sadness	Religion	Social
-All Locations	3.491	1.96	1.521	0.775	0.202	0.308	0.647	5.201
	3.634	2.018	1.491	0.825	0.188	0.131	0.547	4.187
"Arlington, VA"	0.893	0	0.893	0.893	0	0	0	7.011
"Baltimore, MD"	4.35	0	4.35	0	4.35	0	0	0
"Boston, MA"	4.559	3.633	0.926	0	0	0	0	4.456
"Cairo, Egypt / Dubai"	5.904	3.96	1.908	1.096	0.195	0.268	0.355	8.083
"Cairo, Egypt"	5.156	2.799	2.358	1.896	0	0.463	2.381	8.779
"Cambridge, UK"	5.968	2.38	3.588	1.138	1.138	0	0	4.732
"Doha, Qatar"	1.779	1.118	0.727	0.447	0	0.28	0.29	5.394
"Edinburgh, UK"	2.6	1.15	1.45	0	0	1.45	0	5.903
"Edmonton, Alberta, Canada"	0	0	0	0	0	0	0	9.09
"Exeter, UK"	0	0	0	0	0	0	0	0
"Flemington, New Jersey"	0	0	0	0	0	0	0	6.695
"Fucknutsville, DC"	5.88	0	5.88	5.88	0	0	0	5.88
"Goes, The Netherlands"	3.992	2.056	1.936	0.877	0	0.458	0.196	5.981
"Hell's Breakfast Nook, NYC"	4.331	2.345	1.394	0.558	0.309	0.263	0.702	5.884

Fig. 5. Affective Processes Cube for the Egypt Revolt

Using this data, we developed a novel data mining method that attempted to connect tweet sentiment measures and an event (e.g., protest, bombing, etc). Using the aggregated sentiment measures on each day between Jan. 25, 2011 and Feb. 11, 2011, we built classifiers, including libSVM [24], REPTree [25], and IBK [26], to detect the scale of protests for each day. We tested the prediction accuracy of these classifiers using the ground truth obtained from the Timeline of the 2011 Egyptian Revolution [27]. The results show that these classifiers are able to predict whether there were large-scale protests for a day with reasonable performance: libSVM, REPTree, and IBK achieved a prediction accuracy of 83.33%, 83.33%, and 73.33%, respectively. This demonstrates the data mining capability of the text cube approach.

4 Conclusions

Using linguistic features and the text cube approach for the HSCB dimension of sentiment appears to be quite promising. In particular, the use of the text cube provides a useful way to explore and analyze complex data, and in particular to connect language patterns to potential HSCB dimensions. In the present paper we focused on the HSCB dimension of sentiment. While sentiment analysis has received substantial attention in the literature, we believe that it is novel to apply the text cube approach on this dimension.

We have added two new capabilities to the text cube. The first is the heatmap functionality that provides a geographical overlay of the relevant HSCB derived from linguistic data extracted from the Twitter stream. In the present case, we showed how sentiment varied within a region of interest in a US city, highlighting areas that were producing more negative affect than surrounding areas. We believe this kind of visualization tool can provide meaningful data to analysts that are often overwhelmed by the sheer volume of data produced by Twitter feeds. We also believe that this approach will be useful in other geographic regions, as we suggested in our research on the Arab spring.

The second capability is the data mining functionality. The text cube architecture supports the development of prediction models using the data taken from cubes. For example, models that detect events, such as large-scale protests in the Egyptian Revolution, can be built using the sentiment features stored in an event data cube.

Our future research will expand the text cube approach to other HSCB dimensions. We are currently working on group dynamics, leadership, fatalism, and deception, but believe that the analytic community will find value in the text cube approach for many other HSCB dimensions.

Acknowledgement. We thank the reviewers for the valuable comments. Part of this research was funded by Navy Grant # N00014-11-M-0102 awarded to IAI.

References

1. Numrich, S.K., Tolk, A.: Challenges for Human, Social, Cultural, and Behavioral Modeling. SCS M&S Magazine 1(1) (January 2010)

2. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with Twitter: What 140 characters reveal about political sentiment. In: International AAAI Conference on Weblogs and Social Media (2010)
3. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.: Measuring User Influence in Twitter: The Million Follower Fallacy. In: Fourth International AAAI Conference on Weblogs and Social Media (2010)
4. Lerman, K., Ghosh, R.: Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. In: Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, May 23-26 (2010)
5. Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., Pirahesh, H.: Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tab, and Sub Totals. Data Mining and Knowledge Discovery 1(1), 29–53 (1997)
6. Liu, X., Tang, K., Hancock, J., Han, J., Song, M., Xu, R., Manikonda, V., Pokorny, B.: SocialCube: A Text Cube Framework for Analyzing Social Media Data. In: Proceedings of ASE International Conference on Social Informatics, Washington, DC (December 2012)
7. Lin, C., Ding, B., Han, J., Zhu, F., Zhao, B.: Text Cube: Computing IR Measures for Multidimensional Text Database Analysis. In: Proc. 2008 Int. Conf. on Data Mining, Pisa, Italy (December 2008)
8. Zhang, D., Zhai, C., Han, J.: Topic Cube: Topic Modeling for OLAP on Multidimensional Text Databases. In: Proc. 2009 SIAM Int. Conf. on Data Mining, Sparks, NV (April 2009)
9. Zhang, D., Zhai, C., Han, J.: MiTexCube: MicroTextCluster Cube for Online Analysis of Text Cells. In: Proc. 2011 NASA Conf. on Intelligent Data Understanding, Mountain View, CA (October 2011)
10. Zhao, B., Lin, C.X., Ding, B., Han, J.: TEXplorer: Keyword based object ranking and exploration in multidimensional text databases. In: Int. Conf. on Information and Knowledge Management (October 2011)
11. Liu, X., Tang, K., Buhrman, J.R., Cheng, H.: An agent-based framework for collaborative data mining optimization. In: IEEE International Symposium on Collaborative Technologies and Systems (2010)
12. Tang, K., Liu, X., Tang, Y., Manikonda, V., Buhrman, J.R., Cheng, H.: ABMiner: A scalable data mining framework to support human performance analysis. In: International Conference on Applied Human Factors and Ergonomics (July 2010)
13. Brown, C., Frazee, J., Beaver, D., Liu, X., Hoyt, F., Hancock, J.: Evolution of Sentiment in the Libyan Revolution (2011), White Paper at `https://webspace.utexas.edu/dib97/libya-report-10-30-11.pdf`
14. Liu, X., Hancock, J., Zhang, G., Xu, R., Bazarova, N.: Exploring linguistic features for deception detection in unstructured text. In: Hawaii International Conference on System Sciences, January 4-7 (2012)
15. Ekman, P.: Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. Norton & Company Inc., New York (2001)
16. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178 (1980)
17. Mehrabian, A.: Nonverbal communication. Aldine-Atherton, Chicago (1972)
18. Hancock, J.T., Landrigan, C., Silver, C.: Expressing emotion in text. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2007), pp. 929–932 (2007)
19. Hancock, J.T., Gee, K., Ciaciaco, K., Mae, J.: I'm sad you're sad: Emotional contagion in CMC. In: Proceedings of the ACM Conference on Computer-Supported Cooperative Work (2008)

20. Kramer, A.D.I.: An unobtrusive behavioral model of "Gross National Happiness". In: Proceedings of the ACM Conference on Human Factors in Computing Systems (2010)
21. Golder, S., Macy, M.: Diurnal and Seasonal Mood Vary with Work, Sleep and Daylength across Diverse Cultures. Science 333, 1878–1881 (2011)
22. Pennebaker, J.W., Booth, R.J., Francis, M.E.: Linguistic Inquiry and Word Count: LIWC. LIWC, Austin, http://www.liwc.net
23. Schwarz, N., Clore, G.L.: Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States. JPSP 45, 513–523 (1983)
24. Fan, R.-E., Chen, P.-H., Lin, C.-J.: Working set selection using the second order information for training SVM. Journal of Machine Learning Research 6, 1889–1918 (2005)
25. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
26. Aha, D., Kibler, D., Albert, M.: Instance-based learning algorithms. Machine Learning 6, 37–66 (1991)
27. http://en.wikipedia.org/wiki/Timeline_of_the_2011-2012_Egyptian_revolution