Inconsistencies in Big Data

Du Zhang

Department of Computer Science
California State University
Sacramento, CA 95819-6021
zhangd@ecs.csus.edu

Abstract

We are faced with a torrent of data generated and captured in digital form as a result of the advancement of sciences, engineering and technologies, and various social, economical and human activities. This big data phenomenon ushers in a new era where human endeavors and scientific pursuits will be aided by not only human capital, and physical and financial assets, but also data assets. Research issues in big data and big data analysis are embedded in multi-dimensional scientific and technological spaces. In this paper, we first take a close look at the dimensions in big data and big data analysis, and then focus our attention on the issue of inconsistencies in big data and the impact of inconsistencies in big data analysis. We offer classifications of four types of inconsistencies in big data and point out the utility of inconsistency-induced learning as a tool for big data analysis.

***Keywords**: big data, big data analysis, inconsistencies in big data, inconsistency-induced learning.*

1. Introduction

Today, the advancement of sciences, engineering and technologies, the human endeavors, and the social and economic activities have collectively created a torrent of data in digital form. This big data phenomenon will only get intensified and diversified in the years to come. How to turn this big data phenomenon into a positive force for good has drawn tremendous and intensified interest from an ever-increasing set of big data stakeholders. As big data becomes an increasingly popular buzzword, we must not lose sight of the fact that research issues behind big data and big data analysis are embedded in multi-dimensional scientific and technological spaces. We must maintain a clear picture of what big data is and what big data analysis entails.

The objectives of big data analysis are varied. They are largely aligned with the objectives of big data stakeholders. These can translate into creating values in healthcare, accelerating the pace of scientific discoveries for life and physical sciences, improving the productivity in manufacturing, developing a competitive edge for business, retail, or service industries, and innovating in education, media, transportation, or government. How to better utilize data assets, in addition to physical and financial assets, and human capital, to create value has become a fertile ground for enterprises to gain competitive advantages. As big data analysis becomes the next frontier for advancement of knowledge, innovation, and enhanced decision-making process, the significance of its impact on the society as a whole can never be underestimated.

Many domains and economic sectors can benefit from the big data push: life and physical sciences, medicine, education, healthcare, location-based services, manufacturing, retail, communication and media, government, transportation, banking, insurance, financial services, utilities, environment, and energy industry [3, 14].

In this paper, we first take a close look at various dimensions in big data and big data analysis, and highlight some major issues in those dimensions. We then focus our attention on an important issue: inconsistencies in big data and their impact on the outcome of big data analysis. Inconsistencies are ubiquitous in the real world, manifesting themselves in a plethora of human behaviors and decision-making processes for which big data are acquired, integrated, analyzed, and utilized in an attempt to create values and accelerate scientific discoveries [4,12,15-16,18-19,22,23-27]. Once captured in big data, inconsistencies can occur at various granularities of knowledge content, from data, information, knowledge, meta-knowledge, to expertise [27]. If not handled properly, inconsistencies can have adverse impact on the quality of the outcomes in big data analysis process [1,7]. Inconsistencies can also exhibit in reasoning methods, heuristics, or problem-solving approaches of various analysis tasks, creating challenges for big data analysis.

In the paper, we describe classifications for four types of inconsistencies in big data. It turns out that big data inconsistencies can be utilized as important heuristics for improving the performance in various analysis tasks and the quality of outcomes in big data analysis. The recently proposed inconsistency-induced learning, or i^{2}Learning [28-30], offers a promising approach toward proper handling of big data inconsistencies.

Proc. 12th IEEE Int. Conf. on Cognitive Informatics & Cognitive Computing (ICCI*CC'13)
D.F. Hsu, Y. Wang, A.R. Rao, D. Zhang, W. Kinsner, W. Pedrycz, R.C. Berwick & L.A. Zadeh (Eds.)
978-1-4799-0783-0/13/$31.00 ©2013 IEEE

The rest of the paper is organized as follows. Section 2 gives a panoramic view of big data and big data analysis in terms of the issues and challenges. In Section 3, we focus our attention on four types of inconsistencies in big data and how they impact on the big data analysis. Section 4 discusses how inconsistency-induced learning can be utilized as a tool to turn big data inconsistencies into helpful heuristics for better analysis results. Finally, Section 5 concludes the paper with remarks on future work.

2. Dimensions in Big Data

After surveying the landscape, we summarize various dimensions of big data and big data analysis in Figure 1. As a technical term, big data generates many different interpretations and definitions. A meta-definition based on the size dimension is given in [13]: “big data should be defined at any point in time as ‘data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time’.” The volume-variety-velocity definition [11] attempts to capture not only the size dimension, but also the types and speed (at which data are generated) dimensions of the datasets we encounter today. The survey results in [20] indicated a list of alternative definitions for big data. What has been glossed over in the literature is the following: what exactly does a dataset contain, primitive data elements, or pieces of information, knowledge, or meta-knowledge, or any combination of them? The terms of data and information have been used interchangeably in the literature, but there are distinct definitions for data, information, knowledge, meta-knowledge, and expertise, respectively.

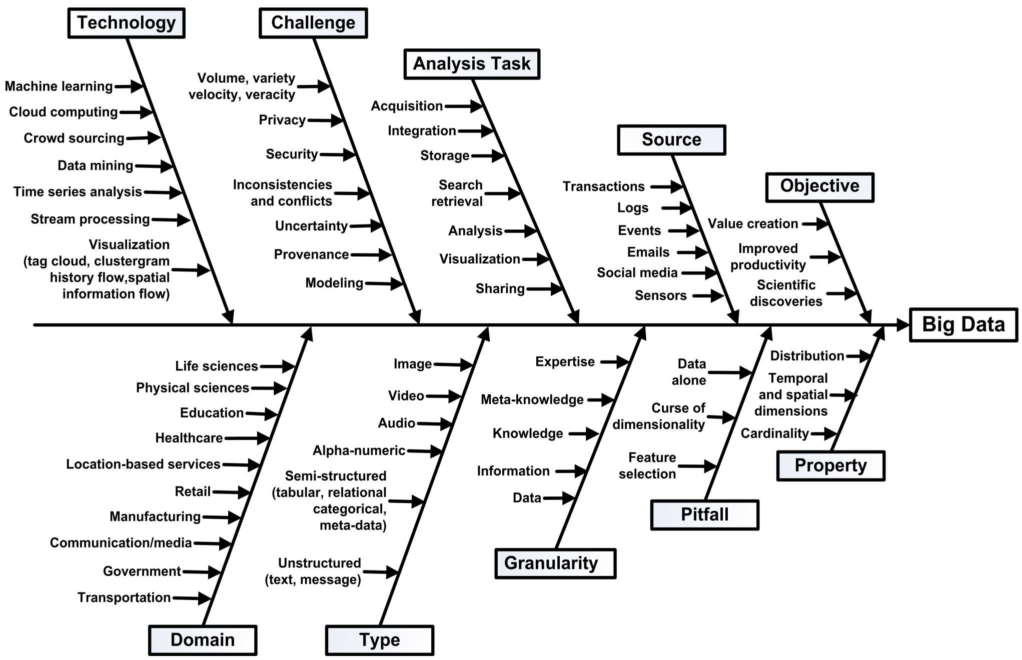

Figure 1. Dimensions in Big Data.

In [1], big data analysis is defined to be a pipeline of acquisition and recording; extraction, cleaning and annotation; integration, aggregation and representation; analysis and modeling; and interpretation. Additionally, there are other alternative definitions on what big data analysis entails [14, 21].

Sources of big data include: transactions, scientific experiments, genomic investigations, logs, events, emails, social media, sensors, RFID scans, texts, geospatial data, audio data, medical records, surveillance, images, and videos [3, 20]. These sources of big data can be semi-structured (e.g., tabular, relational, categorical, or meta-data), or unstructured (e.g., text, messages).

Instances in a dataset can have many properties. For example, data instances may have the same or different probabilistic distributions. As keenly observed in [13], “what makes big data *big* is repeated observations over time and/or space.” Hence, “most large datasets have inherent temporal or spatial dimensions, or both” [13]. Recognizing this inherent temporal/spatial property is very important because this is where performance problems stem from when we try to conduct big data analysis using the prevailing database model (current RDBS model does not honor “the order of rows in tables” [13]). Another property is that most large datasets “exhibit predictable characteristics” in the following sense: the largest cardinalities of most datasets – specifically, the number of distinct entities about which observations are made – are small compared with the total number of observations" [13]. This is a very important heuristic in big data analysis. For scientific datasets, they are typically multi-dimensional, have embedded physical models, possess meta-data about experiments and their provenance, and have low update rates with most updates append-only [2].

The success of big data analysis depends critically on the following array of technologies: machine learning, cloud computing, crowd sourcing [21], data mining, time series analysis, stream processing, and visualization [5, 14].

Big data analysis faces many challenges. In addition to volume, variety and velocity that create challenges in storage, curation, search, retrieval, and visualization issues, veracity generates data uncertainty handling complications [20]. There are a whole host of inconsistent or conflicting circumstances during big data analysis [1, 7]. How to properly handle various types of inconsistencies during data pre-processing and analysis is another challenge. Additional challenges exist that include privacy, security, provenance, and modeling [1, 14].

In the pursuit of advancing knowledge or creating value out of data, there are potential pitfalls. While data are plentiful in today's digital society, we need to be mindful that data alone are not enough to advance knowledge or create value. "Every learner must embody some knowledge or assumptions beyond the data it is given in order to generalize beyond it" [8]. The curse of dimensionality is another potential snag. When utilizing machine learning algorithms to generalize beyond the input data, "generalizing correctly becomes exponentially harder as the dimensionality (number of features) of the examples grows, because a fixed-size training set covers a dwindling fraction of the input space. Even with a moderate dimension of 100 and a huge training set of a trillion examples, the latter covers only a fraction of about 10^{-18} of the input space" [8]. A related issue is feature engineering [8], the large dataset in its raw format "is not in a form that is amenable to learning, but you can construct features from it that are."

In the depth of knowledge, there are layers of knowledge content. Noise can be described as items that carry no content of knowledge. *Data* denotes values drawn from some domain of discourse. *Information* defines the meanings of data values as understood by those who use them. *Knowledge* represents specialized information about some domain that allows one to make decision. *Meta-knowledge* is knowledge about knowledge. *Expertise* is specialized operative knowledge that is inherently task-specific and relatively inflexible. Knowledge content of large granularity has small connotations and knowledge content of small granularity has large connotations. Table 1 summarizes properties and structures in levels of knowledge content in the knowledge hierarchy. Table 2 indicates types of reasoning involved in the knowledge hierarchy: induction that goes from data to knowledge (bottom-up arrow), deduction that applies knowledge to obtain conclusions for data (top-down arrow), and transduction that allows conclusions to be drawn for new data directly from existing data without formulating the knowledge for new data (double arrow).

Table 1. Levels of knowledge content in big data.

Content	**Property**	**Structure**
Expertise	Specialized operative knowledge, task-specific, relatively inflexible	More structured Rich semantics Small connotations Advanced representation
Meta-knowledge	Control strategies Learning decisions	
Knowledge	Declarative Procedural	
Information	Functional dependencies Associations	
Data	Symbolic, numeric, text, graphic, categorical, waveform, video	Less structured Simple semantics Large connotations Simple representation

Table 2. Types of reasoning in big data analysis.

Content	**Decision Process**	
	Inductive	*Deductive*
Expertise		
Meta-knowledge		
Knowledge		
Information		
Data	*Transductive*	

As a categorical phrase, big data has been used to refer to large datasets. But what exactly such a large dataset contains is what we need to be precise and specific about. "Data", "information", and "knowledge" should not be regarded as interchangeable terms denoting the same entities and having the same set of connotations. Table 3 offers some examples illustrating the differences.

We believe that bringing concepts in granularities of knowledge content explicitly into the big data analysis can lead to a better and more accurate curation. It is conducive to various tasks at different stages in the analysis process. For instance, depending on the circumstance of an input set (e.g., containing data elements only, or data elements plus domain knowledge), a learning algorithm that works best under the circumstance can be selected. With regard to the terminology, in addition to big data, terms of *big information*, *big knowledge*, or *big meta-knowledge* can be more pertinently utilized to accurately describe circumstances where an input set contains large volume of information, knowledge, or meta-knowledge, respectively.

Table 3. Examples of knowledge content granularities in big data.

	Location-based services	Social networks	Healthcare	Retail
Knowledge	Restaurant ratings	Social network structures	Diagnoses	Purchase patterns
Information	Restaurants	People who tweet and people who follow other people	Patients	Groups of customers
Data	Latitude-longitude coordinates	Tweets	X-ray images	Transactions

3. Big Data Inconsistencies

In circumstances where big data are produced, acquired, aggregated, transformed, or represented, inconsistencies invariably find their way into large datasets. This can be attributed to a number of factors in human behaviors and in decision-making process. Once captured in big data, inconsistent or conflicting phenomena can occur at various granularities of knowledge content, from data, information, knowledge, meta-knowledge, to expertise, and can adversely affect the quality of the outcomes in big data analysis process [1, 7]. Inconsistencies can also arise in reasoning methods, heuristics, or problem-solving approaches deployed for various analysis tasks, resulting in complications for big data analysis.

Before deciding on how to go about with the inconsistencies found in big data, we need to recognize different types of inconsistencies for different types of big data. For instance, for location-based data, temporal or spatial inconsistencies [27] will dominate, whereas for unstructured text data, inconsistencies stemming from syntactic, semantic, and pragmatic circumstances of a natural language will occupy a commanding position. How we identify and differentiate categories of inconsistent phenomena at levels of data, information, knowledge, meta-knowledge will help pave the way for subsequent handling tasks. Inconsistencies at data level involve various types of values (symbolic, numeric, categorical, waveform, etc.). Inconsistencies at information level manifest in terms of functional dependencies or associations. At knowledge level, inconsistencies display in declarative or procedural beliefs. Meta-knowledge inconsistencies are demonstrated through control strategies or learning decisions [27]. Analysis is needed to establish correspondence between big data analysis tasks and types of inconsistencies impacting or affecting those tasks. Finally, we can use the analysis task to inconsistent phenomena correspondence to guide the development of task-inconsistency specific methods or tools to help assist various tasks in big data analysis. One example is inconsistency-induced learning, or i^{2}Learning in [28-30], that allows inconsistencies to be utilized as stimuli to initiate learning episodes. The outcome of such perpetual learning mechanism will result in refined or augmented knowledge that in turn improves the big data analysis performance.

In the rest of this section, we will elaborate on four important types of inconsistencies in big data. They include: temporal inconsistencies, spatial inconsistencies, inconsistencies found in unstructured natural language text, and inconsistencies stemming from violations of functional dependencies.

3.1. Temporal Inconsistencies

When datasets contain a temporal attribute, data items with conflicting circumstances may coincide or overlap in time. The time interval relationships between conflicting data items can result in partial temporal inconsistency or complete temporal inconsistency [23]. Temporal inconsistencies have been utilized as problem-solving heuristics in IBM Watson open-domain QA system where temporal reasoning is deployed to “detect inconsistencies between dates in the clue and those associated with a candidate answer” [10]. In a temperature time-series data, a temperature recording of 35°F in July in New Orleans would be inconsistent with the context. In human electrocardiogram time-series data, a prolonged period of low value output in the ECG is inconsistent with the normal heart rhythm pattern, an indication for atrial premature contraction [6]. Table 4 gives a list of temporal inconsistencies.

Table 4. Temporal inconsistencies.

	Conflicting case
Partial	Time intervals of two inconsistent events are partially overlapping.
Complete	Time intervals of two inconsistent events coincide or satisfy containment.
Anomalous value	A time-series data has an anomalous value.
Contextual	A time-series data has an anomalous instance in a given context.
Motif	Time-series data has a segment of data values that reoccurs and is anomalous

3.2. Spatial Inconsistencies

When datasets include geometric or spatial dimension, data items are often about objects in space that have geometric properties (location, shape) and that observe spatial relations (topological, directional and distance relations) (see Table 5). Spatial inconsistencies can arise from the geometric representation of objects (a spatial object having multiple conflicting geometric locations), spatial relations between objects (violations of spatial constraints with regard to some spatial relation), or aggregation of composite objects (different representations of the same object from different sources resulting in violation of the constraint that objects must have unique geometric representation) [19].

Table 5. Spatial relations.

	Rotation	Translation	Scaling
Topological	invariant	invariant	invariant
Directional	changing	invariant	invariant
Distance	invariant	invariant	changing

Again in IBM Watson system, geospatial reasoning is deployed that is capable of detecting spatial inconsistencies stemming from conflicting spatial relations such as “directionality, borders, and containment between geo-entities” [10]. Table 6 is a list of possible spatial inconsistencies based on [19].

Table 6. Spatial inconsistencies.

	Conflicting case
Geometric location	A spatial object has conflicting geometric locations.
Geometric shape	A spatial object has conflicting geometric shapes.
Topological	Violation of topological constraints.
Directional	Violation of directional properties.
Distance	Violation of distance properties.
Scaling induced	Different geometric representations of a spatial object at different scales.
Semantic constraint	Violation of semantic integrity constraints.
Structural constraint	Violation of structural integrity constraints of geometric primitives.
Integration induced	Different representations of the same spatial object from different sources resulting in violation of the constraint that objects must have unique geometric representation.

Another example is the work in [22] that defined a concept called conflation that reconciles spatial inconsistencies arising from combining information from diverse sources. The work deals with conflating road maps with aerial images using road intersections as conjugate features and is based on a three-process workflow that includes preprocessing for road candidate identification, spatial inconsistency removal, and shape disagreement removal. The results indicated that the conflated approach yields a 36.6% accuracy improvement over the non-conflated approach for the experiments.

3.3. Text Inconsistencies

As big datasets are increasingly generated from social media, blogs, emails, crowd-sourced ratings, inconsistencies in unstructured text and messages become an important research topic [15,18]. If two texts are referring to the same event or entity, then they are said to be of co-reference [15]. Event or entity co-referencing is a necessary condition for text inconsistencies [15]. Table 7 summarizes a list of text inconsistencies.

Table 7. Text inconsistencies.

	Conflicting case
Complementary	- Miami Heat was in the 2012 NBA final. - Miami Heat was not in the 2012 NBA final.
Mutual exclusive	- Sea cucumber is animal. - Sea cucumber is vegetable.
Inheritance	- Penguin cannot fly. - Penguin is bird and bird can fly, hence penguin can fly.
Synonym	- The system has a fast response time. - The system’s response time is not rapid.
Antonym	- The system has a fast response time. - The system has a slow response time.
Anti-subsumption	- John is a surgeon. - John is not a doctor.
Anti-supertype	- BigDog is not a robot. - BigDog is a legged squad support system.
Asymmetric	- John is married to Jane. - Jane is not married to John.
Anti-inverse	- John is parent of Mike. - Mike is not child of John.
Mismatching	- M_5 is a mobile agent that runs in both Android and iOS environments. - M_5 does not run in Android environment.
Disagreeing	- M_5 has a memory of 10 GB. - M_5 has a memory of 5000 MB.
Contradictory	- M_5 was developed in March 2013. - M_5 was deployed in December 2012.
Precedence	- Obama succeeded Bush as president in 2009. - Bush succeeded Obama as president in 2009.
Factive [15]	- The bombers had not managed to enter the embassy. - The bombers entered the embassy.
Lexical [15]	- John said Jane did nothing wrong. - John accuses Jane.
World knowledge [15]	- Microsoft Israel, one of the first Microsoft branches outside the USA, was founded in 1989. - Microsoft was established in 1989.
Functional [18]	- Mozart was born in Salzburg. - Mozart was born in Vienna.

3.4. Functional Dependency Inconsistencies

Many big datasets are stored, aggregated, and cleaned through the help of relational database systems where functional dependencies (FD) [16] or conditional functional dependencies [9] play a critically important role in enforcing the integrity constraints for the database. Violations of such functional dependencies or conditional functional dependencies will result in inconsistencies in data and information (Table 8) [9,16].

Table 8. Functional dependency inconsistencies.

	Conflicting case
Single FD	Violation of single functional dependency
Multiple FD	Violation of multiple functional dependencies.
Conditional FD	Violation of conditional functional dependencies

3.5. Occurrence in Knowledge Content Levels

The aforementioned types of inconsistencies can manifest themselves at various levels of knowledge content. Table 9 summarizes possible occurrences of various types of inconsistencies at different levels of knowledge content.

Table 9. Occurrence in knowledge content levels.

	Data	Information	Knowledge	Meta-K
Temporal	√	√	√	
Spatial	√	√		
Text	√	√	√	
FD	√	√		

4. Inconsistency-Induced Learning

A framework for inconsistency-induced learning, or i^2Learning, has been proposed in [28-30]. i^2Learning accommodates perpetual or lifelong learning by allowing successive learning episodes to be triggered through inconsistencies an agent encounters during its problem-solving episodes. Learning in the framework is accomplished through the continuous knowledge refinement and/or augmentation in order to overcome encountered inconsistencies. An agent's performance at tasks can be incrementally improved with each learning episode. i^2Learning offers an overarching structure that facilitates the growth and expansion of various inconsistency-specific learning strategies.

The essential idea behind i^2Learning is to identify the cause of inconsistency and then apply cause-specific heuristics to resolve inconsistencies. For instance, if an inconsistent phenomenon stems from irrelevant features, then we can deploy a search algorithm that discerns relevant features from irrelevant ones [29]. We can then overcome inconsistencies by excluding irrelevant features from participating in the analysis process. If an inconsistent case arises as a result of property inheritance, then the heuristic rules of subclass-specificity superseding superclass-generality can be utilized to resolve inheritance inconsistencies [30].

In the context of big data and big data analysis, i^2Learning can also play an active role in improving the data quality by reconciling the inconsistencies found in the datasets, in refining or augmenting knowledge for analysis, modeling or interpretation of big data, and in helping enhance big data applications. For instance, in a crowd-sourced customer ranking application for goods or services, comments made by customers invariably contain inconsistencies (text, temporal, or spatial). Treating customers' comments as a knowledge base that contains pockets of incompatible opinions, we can apply i^2Learning algorithms to resolve or overcome the inconsistencies in customers' comments, which in turn refines or augments this "knowledge base" to render a more coherent and accurate ratings of the goods or services. This process is continuous and perpetual, with each new inconsistent customer comment on things triggering the next learning episode.

5. Conclusion

In this paper, we highlight the multi-dimensional issues and challenges in big data and big data analysis. We then focus our attention on one of the challenges, inconsistencies in big data and their impact on big data analysis. We examine four types of inconsistencies in big data, namely, temporal inconsistencies, spatial inconsistencies, text inconsistencies, and functional dependency inconsistencies. The contribution of this work lies in the fact that articulating explicitly the types of inconsistent phenomena in big data can help pave the way to improve the quality of big data analysis.

Future work can be carried out in the following directions. Details of other frequently encountered types of inconsistencies in big data and their impact on big data analysis still need to be fleshed out. Empirical study is planned on utilizing i^2Learning algorithms with some real world dataset to improve the analysis results.

Acknowledgements. We express our appreciation to four anonymous reviewers for their comments that help improve the content and the presentation of this paper.

References

[1] D. Agrawal, P. Bernstein, E. Bertino, S. Davidson, and U. Dayal, Challenges and opportunities with big data, Cyber Center Technical Report 2011-1, Purdue University, January 1, 2011.

[2] A. Ailamaki, V. Kantere, and D. Dash, Managing scientific data, *Communications of the ACM*, Vol.53, No.6, (2010) 68-78.

[3] Big Data, http://en.wikipedia.org/wiki/Big_data

[4] R.J. Brachman, and H.J. Levesque, *Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers, 2004.

[5] S. Bryson, D. Kenwright, M. Cox, D. Ellsworth, and R. Haimes, Visually exploring gigabyte data sets in real time, *Communications of the ACM*, Vol.42, No.8, (1999) 83-90.

[6] V. Chandola, A. Banerjee, and V. Kumar, Anomaly detection: a survey, *ACM Computing Surveys* 41 (3) 2009, pp.1-58.

[7] S. Chaudhuri, U. Dayal, and V. Narasayya, An overview of business intelligence technology, *Communications of the ACM*, Vol.54, No.8, (2011) 88-98.

[8] P. Domingos, A few useful things to know about machine learning, *Communications of the ACM*, Vol.55, No.10, (2012) 78-87.

[9] W. Fan, F. Geerts, X. Jia, and A. Kementsietsidis, Conditional functional dependencies for capturing data inconsistencies, *ACM Transactions on Database Systems*, Vol. 33, Issue 2, June 2008.

[10] D. Ferrucci, et al, Building Watson: An Overview of the DeepQA Project, *AI Magazine*, Fall 2010, pp.59-79.

[11] Gartner Group press release, Pattern-based strategy: getting value from big data, July 2011.

[12] R. Gotesky, The Uses of Inconsistency. *Philosophy and Phenomenological Research*. Vol. 28, No. 4, 1968, pp.471-500.

[13] A. Jacobs, The pathologies of big data, *Communications of the ACM*, Vol.52, No.8, (2009) 36-44.

[14] J. Manyika, M.Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, Big Data: the next frontier for innovation, competition, and productivity, McKinsey Global Institute, June 2011.

[15] M-C de Marneffe, A. N. Rafferty and C. D. Manning, Finding Contradictions in Text, Proc. of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2008, pp.1039-1047.

[16] M. V. Martinez, A. Pugliese, G. I. Simari, V. S. Subrahmanian, and H. Prade, How dirty is your relational database? An axiomatic approach, in *Proc. 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Hammamet, Tunisia, LNAI 4724, 2007, pp.103-114.

[17] G. Mone, Beyond Hadoop, *Communications of the ACM*, Vol.56, No.1, (2013) 22-24.

[18] A. Ritter, D. Downey, S. Soderland and O. Etzioni, It's a Contradiction-No, It's Not: A Case Study Using Functional Relations, *Proc. of Conference on Empirical Methods in Natural Language Processing*, 2008.

[19] A. Rodriguez, Inconsistency issues in spatial databases, in L. Bertossi et al (eds.) *Inconsistency Tolerance,* LNCS 3300, Springer-Verlag, 2004, pp.237-269.

[20] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, Analytics: the real-world use of big data: how innovative enterprises extract value from uncertain data, Executive Report, IBM Institute for Business Value and Said Business School at the University of Oxford, 2012.

[21] The White House Big Data Research and Development Initiative, http://*www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf*

[22] S. Yang, C. Kim, Y. Huh, and K. Yu, Removal of spatial inconsistency and shape disagreement for conflation of road maps with aerial images, *Canadian Journal of Remote Sensing*, 38(06), 2012 pp.723-737.

[23] D. Zhang, On Temporal Properties of Knowledge Base Inconsistency. *Springer Transactions on Computational Science* V, LNCS 5540, 2009, pp.20-37.

[24] D. Zhang, Toward a classification of antagonistic manifestations of knowledge. *Proc. of Twenty Second International Conference on Tools with Artificial Intelligence*, Arras, France, 2010, pp.375-382.

[25] D. Zhang, Inconsistency: the good, the bad, and the ugly. *International Transactions on Systems Science and Applications*, Vol.6, No.2/3, August 2010, pp.131-145.

[26] D. Zhang, The utility of inconsistencies in information security and digital forensics. In T. Özyer et al (ed.) *Recent Trends in Information Reuse and Integration*, Springer-Verlag, 2011, pp.381-397.

[27] D. Zhang and E. Gregoire, The landscape of inconsistency: a perspective, *International Journal of Semantic Computing*, Vol. 5, No.3, 2011, pp.235-256.

[28] Du Zhang and M. Lu, Inconsistency-induced learning for perpetual learners, *International Journal of Software Science and Computational Intelligence*, Vol.3, No.4, 2011, pp.33-51.

[29] D. Zhang, i^2Learning: perpetual learning through bias shifting, in *Proc. of the 24th International Conference on Software Engineering and Knowledge Engineering*, July 2012, pp. 249-255.

[30] D. Zhang and M. Lu, Learning through Overcoming Inheritance Inconsistencies, in *Proc. of* the 13th *IEEE International Conference on Information Reuse and Integration*, August 2012, pp.201-206.