Journal of Information and Knowledge Management (JIKM) Volume 10 Number 1 (2020)

Uncertain data: Forms, Semantics, Processing Queries

Trieu Minh Nhut Le[1], Thanh Ngoc Thi Nguyen[1], and Dung Hoang Thi Ngoc[1]

[1]*Saigon University, Ho Chi Minh, Viet Nam*
nhuttrieu@sgu.edu.vn,thanhptds@gmail.com,ngocdung.binhtho@gmail.com

Received Date: 2 January 2020
Publish Date: 14 April 2020

Abstract. Data discovery is a broad discipline that comprises a various of techniques for extracting meaningful information. In practical, there are various reasons for uncertain data such as error devices of data collection, data transferring in interference environment, statistic or survey data, predict information and so on. Therefore, studying and discovering meaningful information on uncertain data have been more challenges. Some researches have been studied definitions and methods for uncertain data to solve specified scenarios according to satisfying criteria on semantics. In this paper, we first review and classify types of uncertain data in specific scenario. Secondly, some cases of queries processing to demonstrate the semantics on uncertain data are analysed based on their answers. Thirdly, we review almost probabilities of queries that has been used to evaluate the performing queries process according to semantics of uncertain. Lastly, we mention some problems of semantics and their answers which have been argued in doing uncertain data research.
Keywords: uncertain data, queries processing on uncertain data

1 Introduction

The fourth industrial revolution encompasses areas such as all smart devices which have been invented speedily in our real life. The trend towards automation and data exchange in manufacturing technologies and processes. The explosions of data have been stored and recorded from large number of sources. In relational database system, data are processed with completed value. However, uncertain data is general, finding and scrutinizing these kinds of data is a fundamental problem in exploring the hiding information. Identifying and defining the uncertain data have been played in important role in processing information to obtain the result. Moreover, there are a number of studies (Arefin & Morimoto, 2012) (Khalefa, Mokbel, & Levandoski, 2008) (Sajeev & A., 2016) about semantics of uncertain data focusing on identifying the novel methods and solutions to select the suitable answers.

There are many reasons that lead into uncertain values which make a database incomplete, contain probability information or being presented by fuzzy value. These uncertain data could be collected from error devices, processing of business intelligent, artificial intelligent, decision making, data statistics and so on. For example, sensor devices transferring data in noise environment are typically reasons a number of uncertain data sets which may contain error or incomplete. Facebook analysing through millions of access or click stream logs to predict future user's access. In the stock market, the future stokes prices have been predict by analysing and statistic stock data. Similarly, uncertain data is created through the survey and regression techniques for presenting natural situations. Therefore, uncertain data is unavoidable in database system.

A number of issues on modelling the uncertain data (Sajeev & A., 2016) (Le, Cao, & He, Top-k best probability queries and semantics ranking properties on probabilistic databases, 2013), identifying semantics of uncertain answer (Cormode, Li, & Yi, 2009) (Hua, Pei, Zhang, & Lin, 2008), definition the calculating uncertain value (Zhang, Lin, Pei, & Zhang, 2008) (Cheng, Kalashnikov, & Prabhakar, 2003), and processing uncertain data (Le, Cao, & He, Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples, 2016) (Atallah & Qi, 2009) have been discussed. This paper will discuss about difference issue with uncertainty classification, uncertain data representation, and their corresponding effect on execution or processing uncertain results. Moreover, the evaluation properties of each query have been presented and discussed to prove the semantics answers in each cases of assumption.

This paper is organized as follows: Section 2 will classify and examine the representation of uncertain data and their definitions. In Section 3, we will present a processing of each types of uncertain data. Many scenarios are given for the assuming of processing the uncertain value which is explained and discussed. Section 4 summaries almost specific queries to evaluate uncertainty on data.

2 Forms of Uncertain Data

This section has studied the problems of uncertain data in the literature. The uncertainty of data is classified three main types of uncertainty, the incomplete data, probabilistic data, and fuzzy data.

1. Incomplete data.

First type of uncertain is missing data or incomplete data which is the most problem in different database applications such as data integration, web heterogeneous databases, temporal databases, or is collected from error devices, or poor environment in transfer data and so on. These reasons lead to data missing and error. It means that some values in fields are lost, that is tackling the traditional database processing (Antova, Koch, & Olteanu, 2009) (Soliman, Ilyas, & Ben-David, 2010). For many years, the issues of predicting missing values have been given in database. Missing value negatively influences the quality of database which will lead to high rate of bias results of execution on database. The poor quality of data is a cause to destroy almost systems which is the vital problem in incomplete data.

Example 1: In the incomplete data (Arefin & Morimoto, 2012) (Khalefa, Mokbel, & Levandoski, 2008) (Sajeev & A., 2016), tuples with d-dimensional data with the missing values on some dimensions, such as tuple t_1=(5, 3, 4, -, -) and t_2 = (-, 2, 1, 5, 3), tuple t_1 is missing values on 4^{th} and 5^{th} dimension, t_2 is missing value on 1^{st} dimension. There are a lot of reasons for these missing values in a paper (Miao, Gao, Zheng, Chen, & Cui, 2016) mentions about the rating films of audiences are missing. The audiences only rate the films they have been watched. These data are studied for ranking and top-k queries about the films.

2. Probabilistic data.

The second type of uncertain is probabilistic data which present how likely an event is occurred. For example, in business, decision making of investors has been provided advices with predicting values related to successful or unsuccessful projects (Le, Cao, & He, Top-k best probability queries and semantics ranking properties on probabilistic databases, 2013), for that user's investment, the market data is analysed to predict the future trend with the probability based on the historical statistical data. This is known as a probabilistic data. The probabilistic data is mainly modelling into two types uncertain object model and possible worlds semantic model (Le, Cao, & He, Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples, 2016).

a. Uncertain object model.

In this model, there are many of studies (Tao, Cheng, Xiao, Wang Kay Ngai§, & Prabhakar, 2005) (Reynold & Sunil, 2003) (Pei, Jiang, Lin, & Yuan, 2007) have modeled the uncertain data in a continuous case of probabilistic situation. The uncertain objects databased consists a number of objects, each object is described with multiple values being mutually exclusive. It means that one object could have instances with corresponding to probabilities. This value usually presents continuous cases of probability by using a probabilistic density function (PDF) or a probabilistic mass function (PMF). Uncertain object model (Zhang, Lin, Pei, & Zhang, 2008) is represented based on object-level uncertainty in uncertain objects databases.

Example 2: For uncertain object model, the uncertain object data including the set of moving objects for observing the location, or a set of objects in circular wireless area (Miao, Gao, Zheng, Chen, & Cui, 2016). A meteorology system surveys the humidity, temperatures, ultraviolet index and so on which uses the sensors, transferring data, and data streams. These devices are caused by distance, noise environment, delay etc. The value is estimated by using a probability density function (pdf) such as the pdf of an object (x) moving in area equals one divide area of uncertain region (ur), or ($pdf(x)=1/AREA(ur)$).

b. Possible worlds semantic model.

Another probabilistic data is presented with discrete cases. For this condition, the paper (Semantics of Ranking Queries for Probabilistic data and expected ranks) (Yan & Ng, 2011) has been modeled as a possible worlds semantic model in their studies. With the attribute level, the probabilistic data model contains a set of probabilistic tuples, each tuple is associated with a probability which is distribution information describing a set of possible values. The probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty. Based on probability theories, the probabilistic data model also includes some rules, the exclusive rules, and inclusive rules which used to control logic statements in the probabilistic data. Furthermore, the possible world concept has been introduced to find the semantic results of queries on probabilistic data. The cardinality of all possible worlds is a 2^n.

Example 3: The probabilistic data is presented as a set of probabilistic tuples; sensors are used to detect the animals suspected staying in the detection area of sensors (Hua, Pei, Zhang, & Lin, 2008). There are five dimensions (identify, location, sensor-id, duration (minutes), confident) in each tuple t_3 (‘R3’, ‘A’, ‘S231’, ‘12’, 0.6), and t_4 (‘R4’, ‘E’, ‘S063’, 17, 0.8). In some case, there are two or many devices may detect the presence of object at the same time. They have recorded multi-tuples in data set where only one tuple could be correct. These cases are controlled by the exclusive rules, in another way, it can be used the inclusive rules.

3. Fuzzy data.

The fuzzy data is data considerably the context or conditions instead of fixed value. The altogether values are unclear, lacking meaning, or vague in meaningless situations. The concept of fuzzy is proposed to implement the context or conditions which can be vague, imprecise, or meaningless. The fuzzy logic represent semantics of features or language. The fuzzy statement is defined true in some extent which is represented by scale values with membership functions. Nowadays, researchers usually mathematically present the fuzzy concepts in fuzzy values, fuzzy entities, fuzzy constraints, fuzzy logic, fuzzy attributes, fuzzy relationship, fuzzy sets, fuzzy models, fuzzy aggregation, and so on (Zhang, Lin, Pei, & Zhang, 2008) (Mares, 2007) (˘Skrbi´c & Rackovi´c, 2014).

Example 4: most examples are traffic control systems, popular sense. The fuzzy logic is used and applied to control brakes car based on speed, acceleration / taking pictures based on weather, humidity / laundry machine based on cleaning cycle, rinse and detergent and so on. The fuzzy value are denoted the scale of level such as black (1.0) / dark grey (0.75) / grey (0.5) / light grey(0.25) / white (0.0) (˘Skrbi´c & Rackovi´c, 2014).

3 Processing Queries on Uncertain Data

In this section, we will analyse the process of each uncertain data type. There are many queries are used to implement the uncertain process. In this paper, we review two most important queries which are top-k queries and skyline queries. The main point of these queries is to identify the domination between values, a crucial problem in uncertainty. Top-k queries is a useful tool to select the k best scores of tuples as an answers set on database. The score is represented by a value (a single dimension) or a function (cooperation multi-dimensions) (Sajeev & A., 2016) (Le, Cao, & He, Top-k best probability queries and semantics ranking properties on probabilistic databases, 2013) (Miao, Gao, Zheng, Chen, & Cui, 2016) (Benouaret, Benslimane, Hadjali, & Barhamgi, 2011). Another important query is a skyline, which uses the concept of domination to compare between tuples based on multiple dimensions. It is called tuple t_a dominates tuple t_b, which means all considering dimensions in tuple (t_a) are better than all respective dimensions in tuple (t_b), the result of the skyline query is a set of non-dominated tuples (Arefin & Morimoto, 2012) (Khalefa, Mokbel, & Levandoski, 2008) (Le, Cao, & He, Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples, 2016). Both top-k and skyline queries are considered as fundamental operations in processing uncertain database system. Therefore, many researches have been using these queries to implement their research's contributions in uncertain database.

1. Queries processing on incomplete data.

There are many proposals using suitable techniques for missing data or incomplete data. One of the most common processing of uncertain data requires reprocessing data before any operations are performed. We will present several common solutions that are utilized in queries processing in incomplete data.

Firstly, the main goal of solving the incomplete data is to improve the quality data, it means that the missing or error values could be recovered based on some related information. The prediction information, statistics data, or machine learning are important methods for handling incomplete data. The missing or error values are identified by analysing and mining to recover the right value based on relevant conditions using these methods (Alwan, Ibrahim, Udzir, & Sidi, 2018) (Batista & Monard, 2003) (Grzymala-Busse & Hu, 2000). In data communication, Hamming codes can detect the bits errors and correct bits error without resending data (Abuelyaman & Al-Sehibani, 2008). This technique has been used to correct an error during transmissions data in noisy channel or a wireless capturing a weak signal. In communication, this problem must be solved to transmit information reliably. Moreover, Ali A. et al (Alwan, Ibrahim, Udzir, & Sidi, 2018) processes skyline in incomplete data. They proposed the prediction and classification methods of handling incomplete data. Their approach captured the relationships between the attributes by utilizing the concept of mining feature correlations to generate an Approximate Functional Dependencies. The results of skyline are guaranteed with the estimated precision.

Secondly, the methodology of processing uncertain is ignored or assigned the infinity value to the missing information for further consideration processing. The technique is proposed by redefining the dominance relationship concept between entities. In this way, the goal is to eliminate the missing data tuples which are not affected to the results. The most relevant data items are retrieved exclusively with no missing information. Paper (Khalefa, Mokbel, & Levandoski, 2008) proposed the new definition of dominance and non-transitive dominance relation in incomplete data such as considering some data missing the same dimension between tuples, and then their values continue further considering for selecting the approximate skyline answers. The ISkyline algorithm are designed by using the virtual point and shadow skylines to extend dominance in cyclic relations. Jilu S. and Noorjahan V. A.Paper (Sajeev & A., 2016) reviewed some algorithm to find top-k dominating values from incomplete data. The Skyband based algorithm (ESB) (Sajeev & A., 2016) (Gao, Miao, Miao, Chen, & Li, 2014) is introduced with novel concepts including expired skyline, shadow skyline collecting non-dominated data items of every bucket. The Upper Bound Based (UBB) algorithm calculates each dimension using dominance. The Bit Map Index Guide (BIG) algorithm using the bitmap vertical index and a binning strategy to compress the bitmap storage.

2. Queries processing on probabilistic data.

In probabilistic data, probabilistic theories are usually used and calculated the probabilities of tuples. In this model, the goal of almost of researches is to choose the answer set of top-k or skyline according to semantics in a specific domain.

For the uncertain objects model, the answer set of the top-k and skyline queries are obtained by using the probabilistic density function of each object in the context of continuous variables. Novel definitions have been giving to identify the objects which obtain some features of given domain. The correctness of proposed concepts are mathematically proven by using probabilistic theories, derivative, and integral. Calculating the probability of all objects, the skyline queries are mentioned in processing uncertain object [36] due to the fact that all objects are likelihood of being in skyline results, so all objects have to be calculated the dominated probability using a grid technique. Moreover, the other applied the top-k queries uncertainty object using static/ moving query point (Lian & Chen, 2011). They proposed the probabilistic top-k star queries (PTkS) to reduce the search space by effective pruning rules. The method aimed to retrieve k objects being closest to the query point. PTkS also applied for dynamic case where the query point is moving.

For the possible world semantics model, the processing queries are the selecting set of tuples with probability as answers set which has to satisfy some conditions or scenarios as semantics cases. Moreover, the probabilistic theories are used in the context of discrete random variables in this model. Proposed and proven the methods of selecting the answers set of top-k and skyline queries are investigated and discussed to identify the best solution for some specific content, semantically. There are many research studying uncertainty using the possible world semantic model. Paper (Soliman, Ilyas, & Chang, Top-k Query Processing in Uncertain Databases, 2007) has given the answer of top-k queries, a tuple vectors. This vector has the highest sum values of probabilities in all possible worlds. The k-tuples vector has strictly together in the same possible world. The expectation concept in probabilistic theory have been apply in (Cormode, Li, & Yi, 2009). The approach is to multiply the tuples' score and probability, the highest expected scores tuples are added in the answer set. However, the expected score has a problem about the magnitude of values. So, they proposed the E-rank by using the ranking formula to calculate the new ranking score for removing the magnitude limitation. For answering the skyline queries, the proposed method uses the dominance concepts to select the skyline answers on probabilistic data (Le, Cao, & He, Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples, 2016). The new reasonable definitions and domination concepts of probabilistic tuples have been suggested and proven mathematically using the probabilistic theories. The nearest neighbour best probabilistic skyline algorithm are created to optimise the search space with pruning rules.

3. Queries processing on fuzzy data.

On fuzzy data, the database system has redefined and reproved almost relational algebra operators and relational calculus by concept of fuzzy logic such (selection, projection, union, set different, and etc.), and then the novel result has been obtained with interesting discovery. In this model, the fuzzy theories have been mathematically applied and determined. As the result, these contributions applied effectively and widely in our real life.

The fuzzy information has been applied on processing the data to present the real information in our life. The data web services are automatically taking into account into the user references (Benouaret, Benslimane, Hadjali, & Barhamgi, 2011). The approach is determined the relevance services with the fuzzy constrains (functionality). The rank-order services, which uses a fuzzification of Pareto dominance, are considered to compute the top-k service compositions. The membership functions as a preferring clause can be shared by users to implement service. In addition, top-k queries of fuzzy multi-Join of multi-Tables when search as you type (M.Naveena & S.Sangeetha, 2013), the fuzzy multi-Joint technique uses the multi-table to generate various search criteria at different levels on multi table of database by mapping similar queries, and the top-k queries are tested to outperform the ranking top-k queries in search as you type on multi-tables.

Recently, the fuzzy relational model has been studied and defined. Specifically, the type 1 of fuzzy logic has proposed the fuzzy concept, stores value, some semantic aspects, fuzzy attribute, fuzzy aggregation, fuzzy queries such as join, select etc. Moreover, several fuzzy research have been extended and applied to the type 2 fuzzy into uncertain data (Mares, 2007) (ˇSkrbi´c & Rackovi´c, 2014), which has been tested and proven the improving and correctness of syntactic solution, implemented functionalities, and outperformance of the XML system. Another research studied about applying fuzzy data on clustering methods which obtain the better result on k-mean, spectral clustering (Ayed, Halima, & Alimi, 2014). In the future, the development of the type-2 fuzzy clustering and scalability methods in big data will be attracted many attentions.

The number of solutions of different researches on uncertain data are listed in Table 1. We also mentions the name of new definitions and concepts for domination of top-k and skyline queries which is described in this paper.

Table 1. Summary of definitions and processing queries on uncertain data

	Identifying			Processing				Contributions
	Incomplete data	Probabilistic data	Fuzzy data	Top-k queries	Skyline queries	New definition	New algorithms	Problems, definitions, techniques, and algorithms
(Arefin & Morimoto, 2012) (Khalefa, Mokbel, & Levandoski, 2008)	✓				✓		✓	Missing → reprocessing data and skyline set computation based on replacement-bucket technique → compute skyline.
(Alwan, Ibrahim, Udzir, & Sidi, 2018)	✓				✓		✓	Missing → estimated value by mining attributes to generate Approximate Functional Dependencies (AFDs) → retrieved skyline.
(Miao, Gao, Zheng, Chen, & Cui, 2016)	✓			✓		✓	✓	Incomplete → define dominance relationship → extended Skyband + Upper Bound Algorithms → k-objects dominating the maximum number of objects in data set.
(Hua, Pei, Zhang, & Lin, 2008)		✓		✓			✓	Probability (Possible worlds) → semantics probability based on threshold – threshold top-k

								approach with effective pruning rules → answer tuples with probability greater than threshold.
(Atallah & Qi, 2009)		✓			✓	✓	✓	Probability (uncertain data) + threshold → dominance and probabilistic skyline → grid method + weighted dominance counting method → skyline probabilities all the data items.
(Lian & Chen, 2011)		✓		✓		✓	✓	Probability (uncertain data) → define star instance / star object → static PTkS query applied pruning heuristics → set of top-k objects.
(Liu, Yang, Ye, & Lee, 2013)		✓			✓	✓	✓	Probability (Possible worlds) → U-skyline definition → U-Skyline processing using search algorithm with dynamic program + pruning rules → non-dominated tuples vector.
(Ilaria, Paolo, & Marco, 2014)		✓		✓	✓	✓	✓	Probability (Possible worlds) → probabilistic ranking semantics / P-domination and Skyline → skyline queries with (tight/loose) integration rules → skyline set / top-1.
(Le, Cao, & He, Top-k best probability queries and semantics ranking properties on probabilistic databases, 2013)		✓		✓		✓	✓	Probability (Possible worlds) → semantic top-k answers removing threshold/ semantic properties → top-k bestpro algorithm → top-k tuples with highest probability.
								Probability (Possible

(Le, Cao, & He, Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples, 2016)		✓			✓	✓	✓	worlds) → probabilistic non-dominating tuples in skyline set/ calculating skyline probabilities → bestpro-skyline algorithms → results is a non-dominating tuples with best probability.
(Benouaret, Benslimane, Hadjali, & Barhamgi, 2011)			✓	✓		✓	✓	Fuzzy (users preferences) → define fuzzy dominating score → top-k compositions algorithm → top-k compositions (retained services) with highest scores.
(M.Naveena & S.Sangeetha, 2013)			✓	✓			✓	Key word search engine (A search as you type) → Fuzzy multi-Join technique to support multi-Tables, and top-k query search → top-k result with user-specified function.

4 Conclusions

In this work, the main goal is to survey possible types of uncertain data and to summarise major techniques and methods of top-k and skyline queries processing on uncertain data. Firstly, we present the incomplete data, probabilistic data and fuzzy data. Incomplete data, probabilistic data, fuzzy data are mentioned and listed almost relevant researches. After that, the top-k and skyline queries processing have been analysed on each type of uncertainty. We have finalised several researches to recapitulate that the semantics of uncertain data being the key of novel solutions and methods in real-life applications.

This paper tries to generate and highlight some important problems which have been discussed in previous research. The top-k and skyline queries are very helpful and important in numerous of real applications. There are numerous methods being brief reviewing and discussion for the future issue. This paper has just listed some major problems, which can help new researchers to have a general idea and context of uncertainty, especially, define and processing the uncertain data.

References

Arefin, M. S., & Morimoto, Y. (2012). Skyline Sets Queries for Incomplete Data. *International Journal of Computer Science & Information Technology (IJCSIT) , 4* (5), 67-80.

Khalefa, M. E., Mokbel, M. F., & Levandoski, J. J. (2008). Skyline Query Processing for Incomplete Data. *2008 IEEE 24th International Conference on Data Engineering.*

Sajeev, J., & A., N. V. (2016). Top-K Dominating Queries On Incomplete Data : A Survey. *International Journal of Scientific Research in Science and Technology , 2* (6), 359-361.

Le, T. M., Cao, J., & He, Z. (2013, April 24). Top-k best probability queries and semantics ranking properties on probabilistic databases. *Data & Knowledge Engineering* , pp. 248-266.

Cormode, G., Li, F., & Yi, K. (2009). Semantics of Ranking Queries for Probabilistic Data and expected ranks. *International Conference on Data Engineering*, (pp. 305-316).

Hua, M., Pei, J., Zhang, W., & Lin, X. (2008). Ranking Queries on Uncertain Data: A Probabilistic threshold approach. *SIGMOD* , 673-686.

Zhang, W., Lin, X., Pei, J., & Zhang, Y. (2008). Managing Uncertain Data: Probabilistic Approaches. *The Ninth International Conference on Web-Age Information Management.*

Cheng, R., Kalashnikov, D. V., & Prabhakar, S. (2003). Evaluating Probabilistic Queries over Imprecise Data. *Proceedings of the ACM SIGMOD International Conference on Management of Data* , 551-562.

Le, T. M., Cao, J., & He, Z. (2016). Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples. *Preprint submitted to Information Science Journal* , 1-32.

Atallah, M. J., & Qi, Y. (2009). Computing All Skyline Probabilities for Uncertain Data. *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* , 279-287.

Antova, L., Koch, C., & Olteanu, D. (2009). 10ˆ(10ˆ6) Worlds and Beyond: Efficient Representation and Processing. *The International Journal on Very Large Data Bases , 18* (5), 1021-1040.

Soliman, M. A., Ilyas, I. F., & Ben-David, S. (2010). Supporting ranking queries on uncertain and incomplete data. *The very Large Database Journal , 19* (4), 477-501.

Miao, X., Gao, Y., Zheng, B., Chen, G., & Cui, H. (2016). Top-k dominating queries on Incomplete Data. *IEEE Transactions on Knowledge and Data Engineering , 28* (1), 252-266.

Tao, Y., Cheng, R., Xiao, X., Wang Kay Ngai§, B. K., & Prabhakar, S. (2005). Indexing Multi-Dimensional Uncertain Data with Arbitrary Probability Density Functions. *Proceedings of the 31st VLDB Conference*, (pp. 922-933). Trondheim, Norway.

Reynold, C., & Sunil, P. (2003). Managing uncertainty in sensor database. *SIGMOD.*

Pei, J., Jiang, B., Lin, X., & Yuan, Y. (2007). Probabilistic Skylines on Uncertain Data. *VLDB* , 15-26.

Semantics of Ranking Queries for Probabilistic data and expected ranks. (n.d.).

Yan, D., & Ng, W. (2011). Robust Ranking of Uncertain Data. *International Conference on Database Systems for Advanced Applications.*

Mares, M. (2007). Fuzzy Data in Statistics. *Kybernetika.*

ˇSkrbi´c, S., & Rackovi´c, M. (2014). *Fuzzy Databases - Monograph -.*

Benouaret, K., Benslimane, D., Hadjali, A., & Barhamgi, M. (2011). Top-k Web Service Compositions using Fuzzy Dominance Relationship. *IEEE International Conference on Services Computing.*

Alwan, A., Ibrahim, H., Udzir, N., & Sidi, F. (2018). Missing Values Estimation for Skylines in Incomplete Database. *The International Arab Journal of Information Technology , 15* (1), 66-75.

Batista, G. E., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence Journal , 17* (5), 519-533.

Grzymala-Busse, J. W., & Hu, M. (2000). A Comparison of Several Approaches to Missing Attribute Values in Data Mining. *in Porcessing of the second International Conference on Rough Sets and Current Trend in Computing.*

Abuelyaman, E. S., & Al-Sehibani, A.-A. S. (2008). Optimization of the Hamming Code for Error Prone Media. *International Journal of Computer Science and Network Security , 8* (3), 278-285.

Gao, Y., Miao, X., Miao, X., Chen, H. C., & Li, Q. (2014). Processing k-skyband, constrained skyline, and group-by skyline queries on incomplete data. *International Journal of Expert System with Application , 41* (10), 4959-4974.

Lian, X., & Chen, L. (2011). Shooting top-k stars in uncertain databases. *The VLDB Journal* , 20:819-840.

Soliman, M. A., Ilyas, I. F., & Chang, K. C.-C. (2007). Top-k Query Processing in Uncertain Databases. *2007 IEEE 23rd International Conference on Data Engineering.*

M.Naveena, & S.Sangeetha. (2013). Fuzzy Multi-Join and Top-K Query Model for search- as-you- type in Multiple tables. *International Journal of Computer Science and Mobile Computing , 2* (12), 114-118.

Ayed, A. B., Halima, M. B., & Alimi, A. M. (2014). Survey on clustering methods: Towards fuzzy clustering for big data. *2014 6th International Conference of Soft Computing and Pattern Recognition* .

Liu, X., Yang, D.-N., Ye, M., & Lee, W.-C. (2013). U-Skyline: A New Skyline Query for uncertain databases. *IEEE Transactions on Knowledge and Data Engineering , 25*, 945-960.

Ilaria, B., Paolo, C., & Marco, P. (2014). Domination in the Probabilistic World: Computing Skylines for arbitrary correlations and ranking semantics. *ACM Transactions on Database Systems , 39*, 14:1-14:45.