ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 5; Issue 3; May-June: 2017; Page No. 183-189
Journal Approved by UGC

A SURVEY ON TRUTH DISCOVERY METHODS FOR BIG DATA

Mr. Bastin thiyagaraj[1], Dr. A. aloysius[2]

[1]Department of Information Technology, St. Joseph's College (Autonomous), Trichy – 620002, TamilNadu, India.

bastinstar@gmail.com

[2]Department of Computer Science, St. Joseph's College(Autonomous), Trichy-620002, TamilNadu, India.

aloysius1972@gmail.com

Abstract

Increasingly large numbers of embedded Smart phones, Sensors, PCs, Tablets, Computers connected to network, internet Medical data, Business transactions, Data are captured by sensors, Social media/networks, Banking, Marketing, Government data, etc are generating enormous amounts of unstructured data. This data creates new opportunities to extract more value for the areas for which it is needed. Recently, the Big Data is a challenging one by a dramatic increase of data from the physical world. One important property of Big Data is its wide variety, i.e., data about the same object can be obtained from various sources. Most of the time sources provide conflicted data for the same object. It is the challenging one to identify the "True Information" from the noisy information. To overcome such difficulties, Truth discovery methods are developed by estimating weight of the each source that is reliability of the sources. This survey focuses on the methods which are used to find out the true information from the conflicted data and comparisons of methods are used to select the appropriate method based on the types of data.

Key Words: *Truth Discovery, Big data, Jaccard distance, Levenshtein distance*

1. Introduction

Big data is a new term, new-buzz word refers to the explosion of available information and massive amounts of very high-dimensional or unstructured data. According to the Definition of Gartner "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making"[30]. The huge amount of unstructured data is generated from Smart devices, Medical data, Business transactions, Banking, Marketing, Government data, etc. It is the challenging one to extract true data from huge amount of variety of unstructured data. Truth discovery plays vital role in the information age to identify the true data. On one hand it is an important role to find the accurate information more than ever, but on the other hand inconsistent information's have been generated due to the "variety " feature of big data. The truth discovery approaches are benefited in many applications in different fields. Examples include healthcare [1], crowd/social sensing [2,3,4,5,6,7], crowd sourcing [8,9,10], Information extraction [11,12], knowledge graph construction [13,14] and so on. But here the critical decisions are taken based on the reliable information extracted from diverse sources. The general approach voting/averaging treats all information sources equally, so that the exact information cannot be identified. The truth discovery aims to infer source reliability degrees, by which trustworthy information can be discovered.

2. GENERAL PRINCIPLE:

The general principle of truth discovery approaches is to describe three popular ways to model it in practice. However, the gathered information about the same object from various sources may conflict with each other due to errors, missing records, typos, faults, misprinted data, out-of-date data, etc. For example, the search results given by Google for the query like —the height of Mount Everest" include —29,035 feet", —29,002 feet" and —29,029feet"[15]. Among these noisy information, which one is more trustworthy, or

Corresponding author: Mr. Bastin thiyagaraj

which represents the true fact?. In this and many more similar problems, it is important to combine and collect noisy information about the same set of objects or events gathered from various sources to get true and accurate facts [15]. The most important feature of truth discovery is to estimate source reliabilities. To identify the trustworthy information (truths), weighted aggregation of the multi-source data is performed based on the estimated source reliabilities. As both source reliabilities and truths are unknown, the general principle of truth discovery works as follows: If a source provides trustworthy information frequently, it will be assigned a high reliability; meanwhile, if one piece of information is supported by sources with high reliabilities, it will have big chance to be selected as truth. Based on these principles the following methods are given in relating to truth discovery methods.

Iterative methods

In iterative methods [31] the truth computation and source reliability estimation of the truth discovery method depend on each other. In iterative procedures, the truth computation step and source weight estimation step are iteratively conducted until convergence. In the truth computation step, sources weights are sumed to be fixed. Then the truths can be inferred through weighted aggregation such as weighted voting. Each candidate value v receives the votes from sources in the following way:

$$vote(v) = \left(\sum_{s \in \mathcal{S}_v} \frac{w_s}{|\mathcal{V}_s|} \right)^{1.2}$$

where S_v is the set of sources that provide this candidate value, and $|V_s|$ is the number of claims made by source S.

In truth computation step, the sources invest their reliabilities among claimed values, and now, in source weight computation step, they collect credits back from the identified truths as follows

$$w_s = \sum_{v \in \mathcal{V}_s} \left(vote(v) \cdot \frac{w_s/|\mathcal{V}_s|}{\sum_{s' \in \mathcal{S}_v} w_{s'}/|\mathcal{V}_{s'}|} \right)$$

Optimization based methods

The general principle of truth discovery is for the optimization formulation:

$$\arg\min_{\{w_s\},\{v_o^*\}} \sum_{o \in \mathcal{O}} \sum_{s \in \mathcal{S}} w_s \cdot d(v_o^s, v_o^*)$$

Where,***d*** is the distance function, measures the difference between the information of the sources ***(v_o^s),*** and identified truths***(v_o^*).*** d depends on data type of ***m***-th property(39).

For **categorical data** [32], we use ***0-1*** function

$$d_m(v_{im}^{(*)}, v_{im}^{(k)}) = \begin{cases} 1, & if\ v_{im}^{(*)} \neq v_{im}^{(k)} \\ 0, & otherwise \end{cases}$$

For **continuous data** (32), we choose normalized absolute deviation.

$$d_m(v_{im}^{(*)}, v_{im}^{(k)}) = \frac{|v_{im}^{(*)} - v_{im}^{(k)}|}{std(v_{im}^{(1)}, \cdots, v_{im}^{(k)})}$$

For **multi-value data**[32], Jaccard distance is applied.

$$d_m(v_{im}^{(*)}, v_{im}^{(k)}) = \frac{|v_{im}^{(*)} \cup v_{im}^{(k)}| - |v_{im}^{(*)} \cap v_{im}^{(k)}|}{|v_{im}^{(*)} \cup v_{im}^{(k)}|}$$

For **text data**, Levenshtein distance is applied.

In **[16] Levenshtein distance** is a string metric for measuring the difference between two sequences. Informally, the Levenshtein distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other. Levenshtein distance may also be referred to as edit distance.

Mathematically, the Levenshtein distance between two strings ***a,b*** (of length ***|a| |b|*** respectively) is given by ***$lev_{a,b}$ (|a|,|b|)***

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

where $\mathbf{1}(a_i \neq b_j)$ is the indicator function equal to ***0*** when $a_i = b_j$ and equal to ***1*** otherwise, and ***$lev_{a,b}$ (i, j)*** is the distance between the ***i*** characters of ***a*** and the first ***j*** characters of ***b***.

For **graph data** [17], we can use graph edit distance. Graph Edit Distance (GED) is a measure of similarity (or dissimilarity) between two graphs. The Graph

Edit Distance between two graphs g_1 and g_2, written as $GED_{(g1, g2)}$ can defined as

$$GED(g_1, g_2) = \min_{(e_1,\ldots,e_k)\in\mathcal{P}(g_1,g_2)} \sum_{i=1}^{k} c(e_i)$$

where $P_{(g1, g2)}$ denotes the set of edit paths transforming ***g1*** into (a graph isomorphic to) ***g2*** and **c(e)≥0** is the cost of each graph edit operation **e**.

Probabilistic graphical model based methods

There are some truth discovery methods [18, 19] based on the PGMs Probabilistic graphical Model. The past methods are used to find the truth for categorical input, such as authors of books or cast members of movies, and weather data. To address the issues of truth finding on numerical data, a Bayesian probabilistic model or the Gaussian Truth Model (GTM), this can leverage the characteristics of numerical data in a principled manner, and infer the real-valued truth and source quality without any supervision. The general principle of PGM is given as follows

$$\prod_{s\in\mathcal{S}} p(w_s|\beta) \prod_{o\in\mathcal{O}} \left(p(v_o^*|\alpha) \prod_{s\in\mathcal{S}} p(v_o^s|v_o^*, w_s) \right)$$

In this model, each claimed value v_o^s is generated based on the corresponding truth v^*_o and source weight w_s, and function $p(v_o^s \mid v^*_o, w_s)$ links them together.

3. TRUTH DISCOVERY METHODS:

TruthFinder: In [20] the method TRUTHFINDER studies the interaction between websites and the facts they provide and infers the trustworthiness of websites and confidence of facts from each other. If a website is said to be a trustworthy, website has to provide many pieces of true information. Bayesian analysis is adopted to iteratively estimate source reliabilities and identify truths.

AccuSim[17,22]: ACCUSIM augments ACCUPR by considering also value similarity in the same way as TRUTHFINDER does. And used Bayesian models that discover copiers by analyzing values shared between sources

AccuCopy [17,22]**:** ACCUCOPY augments ACCUFORMAT by considering the copying relationships between the sources and weighting the vote count from a source s by the probability that s provides the particular value independently.

2-Estimates [23]**:** This method explores single truth identification of each object, and adopts complementary votes. So that, 2-Estimates use two estimators for the truth of facts and the error of views.

3-Estimates [24]**:** This method augments 2-Estimates, and estimates how hard each fact.

Investment [24]: This approach implies that sources invest their reliability among its claimed values, and confidence of the sources estimated by non linear function of claimed values defined by the sum of invested reliabilities from its providers. Since claims with higher-trust sources get higher belief, these claims become relatively more believed and their sources become more trusted.

SSTF [25]: Semi-supervised truth discovery approach, a small set of labelled truths are used to identify the source reliability estimation. Both mutual exclusivity and mutual supports are adopted to capture the relations among claimed values.

LTM [19]: Latent Truth Model (LTM) which leverages a generative error process. Latent Truth Model is a probabilistic graphical model which works with two types of errors under the scenarios of multiple truths: false positive and false negative. This logic is used LTM to break source reliability into two parameters, one for false positive error and the other for false negative error.

GTM [18]: Gaussian Truth Model is a Bayesian probabilistic approach and designed for solving truth discovery problems on continuous data.

Regular EM [6]: Expectation Maximization (EM) is a general optimization technique for finding the maximum likelihood estimation of parameters in a statistic model where the data are "incomplete" [26]. Regular EM is proposed for crowd/social sensing applications, in which the observations provided by humans can be modelled as binary variables.

LCA [27]: Latent Credibility Analysis (LCA) method, a set of latent parameters are used to model source reliability, and gives more informative source reliabilities to end-users. LCA models can outperform the best fact- finders in both unsupervised and semi-supervised settings.

Apollo-social [28]: Apollo-social method collects the information from users on social media platforms such as Twitter. In social network, a claim made by a user can either be originally published by him or be re-tweeted from other users.

CRH [29]: Conflict Resolution on Heterogeneous (CRH) methods provided the framework to find the truths (or correct information) from multiple conflicting sources. This framework worked with the heterogeneity of data. So that, different types of distance functions used to capture the characteristics of different data types, and the estimation of source reliability is jointly performed.

CATD [21]: Confidence Aware Truth Discovery, the authors derive the confidence interval for the reliability estimation and motivated by the phenomenon that many sources only provide very few observations. It is not reasonable to give a point estimator for source reliability.

4. RELATED WORKS:

The World Wide Web plays an important role for the most people to provide the quality of information. But we could not confirm that all the web sites are providing Quality information. Machine learning approaches are used for distinguishing high-quality and low-quality web pages, where the quality is defined by human preference. Some of the websites provide conflict information. Here the following areas providing the information to find the truth information from the conflict information.

X. L. Dong, et al [21]**, X. Li, et al** [22] they found "how to improve truth discovery? by detecting dependence between sources and analyzing accuracy of sources". Bayesian models were used to discover copiers by analyzing values shared between sources. The truthfulness of Deep Web data is identified in two areas are *Stock (Stock prices*) and *Flight (including scheduled departure/arrival time, actual departure/arrival time, and departure/arrival gate)*. They observed a large amount of inconsistency on data from different sources and also some sources with quite low accuracy. The data fusion methods are applied to reduce the conflicts and finding the truth, analyzed their strengths and limitations, and suggested promising research directions.

A. Galland, et al [23], Single truth identification method is explored. The problems are taken "The birth place of historian" and/or "The capital of Cities". For this type of queries sources may provide complicated answer also. The algorithms Cosine, 2-Estimates and 3-Estimates are experimentally used in synthetic and real world data to identify truth information. 3 – Estimates predicted better results compare with other two algorithms.

J. Pasternack, et al[24] In this paper the truth information identified by experimentally over three domains are city population, basic biographies, and American vs. British spelling, with four datasets. All domains are using the methods VoteDistance cost function and Vote Loss vote redistribution. They experimentally conducted using the data "City population of U.S and American Vs British spelling". They incorporated Priori knowledge with Fact Finding Algorithms (Investment, Pooled Investment, and Average·Log). Priori Knowledge was most dramatic in spelling domain. They found that while prior knowledge is helpful in reducing error, when the user's viewpoint disagrees with the norm it becomes absolutely essential and, formulated as a linear program.

Normally the information provided by the web sources are not quality always as well as trustworthy. Various sources provide different quality of data and often provide inaccurate data and conflict information.

D. Yu, H. Huang, et al [12] they used Semi Supervised Learning method to find the truth value with the help of Ground Truth value. The test their approach SSTF on six real data set containing Book authors. The goal of semi supervised truth discovery is to assign a confidence score to each fact, so that true facts have higher scores than false facts. The method is called Semi-Supervised Truth Finder, or SSTF. They define a confidence score to be a real value between −1 and 1. A score close to 1 indicates we are very confident that a fact is true. A score close to −1 indicates the reverse. A score close to 0 indicates that we do not know if a fact is true or false. Each ground truth fact has a confidence score of 1. Small amount of ground truth value can help greatly to identify trustworthy data sources, and used optimal solution algorithm instead of Iteration.

D. Wang, et al [6] They presented a method of Maximum Likelihood to find out the truth discovery

from "social Sensing data". The data are collected from the human population in terms of "Crowd Sourcing". Examples included cell-phone accelerometers, cameras, GPS devices, smart power meters, and interactive game consoles. The challenges were that in social sensing to ascertain the correctness of data from human population data. Regular EM was proposed for crowd/social sensing applications. They found that "non-trivial estimation accuracy improvements can be achieved by the proposed maximum likelihood estimation approach compared to other state of the art solutions".

D. Wang, et al [28] In this paper they explained that human as a " sensor network" that was identified Twitter resources were the human resources. Here humans provided the information voluntarily or re-tweet the information. And they formulated human sensing problem into three related research questions that i) how human sources considered as participatory sensor? ii) how to filter bad data? And iii) how to use the identified algorithm to sense the correct information from the bad information. A tool called Apollo that used Twitter as a "sensor network" for observing events in the physical world. **Q. Li, et al** [29], Heterogeneity of data has been used to identify the truth by resolving the conflict. Normally truth identification involves only on Categorical or continuous types of data. Many truth discovery approaches have been proposed to identify the truth without any supervision. However, the previous approaches are mainly designed for single-type data and they do not take advantage of a joint inference on data with heterogeneous types. But here in (29) they proposed to resolve conflicts from multiple sources of heterogeneous data types. The objective is to minimize the overall weighted deviation between the truths and the multi-source observations where each source is weighted by its reliability. Conflict Resolution on Heterogeneous Data (CRH) framework can be developed to identify the truth from the conflicting sources of data. Different loss function used along with CRH to identify the characteristics of data. The experiments conducted on real-world weather, stock and flight data as well as simulated multi-source data. They formulated the problem as an optimization problem to minimize the overall weighted deviation between the identified truths and the input.

5. CONCLUSION:

Truth discovery methods are playing an important role in Big data. It is the challenging one to extract true information from the conflicted information and also it is crucial to extract knowledge from the amount of information generated by the various sources. It is very difficult to identify trustworthy information from the multiple conflicting data sources. In this survey paper, the general principal of the truth discovery methods are provided, which are used to find out the deviation from the conflicted data collected from the various sources. The comparison of the various methods are discussed, which are used to select the appropriate method based on the data and to find out the true information based on the types of data (numerical, categorical and continuous).

REFERENCES:

1. S. Mukherjee, G. Weikum, and C. Danescu-NiculescuMizil. People on drugs: credibility of user statements in health communities. *In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14),* pages 65-74, 2014.
2. C. C. Aggarwal and T. Abdelzaher. Social Sensing. *In Managing and mining sensor data*, pages 237-297. 2013.
3. H. Le, D. Wang, H. Ahmadi, Y. S. Uddin, B. Szymanski, R. Ganti, and T. Abdelzaher. Demo: Distilling likely truth from noisy streaming data with apollo. *In Proc. of the ACM International Conference on Embedded Networked Sensor Systems (Sensys'11),* pages 417-418, 2011.
4. C. Miao, W. Jiang, L. Su, Y. Li, S. Guo, Z. Qin, H. Xiao, J. Gao, and K. Ren. Cloud-enabled privacy preserving truth discovery in crowd sensing systems. *In Proc. of the ACM International Conference on Embedded Networked Sensor Systems (Sensys'15),* 2015.
5. L. Su, Q. Li, S. Hu, S. Wang, J. Gao, H. Liu, T. Abdelzaher, J. Han, X. Liu, Y. Gao, and L. Kaplan. Generalized decision aggregation in distributed sensing systems. *In Proc. of the IEEE Real-Time Systems Symposium (RTSS'14),* pages 1-10, 2014.
6. D. Wang, L. Kaplan, H. Le, and T. Abdelzaher. On truth discovery in social sensing: A maximum likelihood estimation approach. *In Proc. of the*

International Conference on Information Processing in Sensor Networks (IPSN'12), pages 233- 244, 2012.

7. S. Wang, L. Su, S. Li, S. Yao, S. Hu, L. Kaplan, T. Amin, T. Abdelzaher, and W. Hongwei. Scalable social sensing of interdependent phenomena. *In Proc. of the International Conference on Information Processing in Sensor Networks (IPSN'15),* pages 202- 213, 2015.
8. B. Aydin, Y. Yilmaz, Y. Li, Q. Li, J. Gao, and M. Demirbas. Crowd sourcing for multiple-choice question answering. *In Proc. of the Conference on Innovative Applications of Artificial Intelligence (IAAI'14),* pages 2946-2953, 2014.
9. H. Li, B. Zhao, and A. Fuxman. The wisdom of minority: discovering and targeting the right group of workers for crowdsourcing. *In Proc. of the International Conference on World Wide Web (WWW'14),* pages 165-176, 2014.
10. J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan. Whose vote should count more: Optimal integration of labelers of unknown expertise. In Advances in Neural Information Processing Systems (NIPS'09), pages 2035-2043, 2009.
11. F. Li, M. L. Lee, and W. Hsu. Entity profiling with varying source reliabilities. *In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14),* pages 1146-1155, 2014.
12. D. Yu, H. Huang, T. Cassidy, H. Ji, C. Wang, S. Zhi, J. Han, C. Voss, and M. Magdon-Ismail. The wisdom of minority: Unsupervised slot filling validation based on multi-dimensional truth-finding. *In Proc. of the International Conference on Computational Linguistics (COLING'14)*, 2014.
13. X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. *In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14),* pages 601- 610, 2014.
14. X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, and W. Zhang. From data fusion to knowledge fusion. PVLDB, 7(10):881-892, 2014.
15. Arunima Kumari, Dr. Dinesh Singh, Reviewing Truth Discovery Approaches And Methods For Big Data Integration. *International Journal of Science, Engineering and Technology Research (IJSETR),* Volume 5, Issue 8, Pages: 2766-2774, August 2016.
16. https://en.wikipedia.org/wiki/Levenshtein_distance
17. https://en.wikipedia.org/wiki/Graph_edit_distance
18. B. Zhao and J. Han. A probabilistic model for estimating real-valued truth from conflicting sources. *In Proc. of the VLDB workshop on Quality in Databases (QDB'12),* 2012.
19. B. Zhao, B. I. P. Rubinstein, J. Gemmell, and J. Han. *A bayesian approach to discovering truth from conflicting sources for data integration. PVLDB,* 5(6):550-561, 2012.
20. X. Yin, J. Han, and P. S. Yu. Truth discovery with multiple conflicting information providers on the web. *In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07),* pages 1048-1052, 2007.
21. X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: The role of source dependence. PVLDB, 2(1):550-561, 2009.
22. X. Li, X. L. Dong, K. B. Lyons, W. Meng, and D. Srivastava. Truth finding on the deep web: Is the problem solved? PVLDB, 6(2):97-108, 2012.
23. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating information from disagreeing views. *In Proc. of the ACM International Conference on Web Search and Data Mining (WSDM'10),* pages 131-140, 2010.
24. J. Pasternack and D. Roth. Knowing what to believe (when you already know something). *In Proc. of the International Conference on Computational Linguistics (COLING'10)*, pages 877-885, 2010.
25. X. Yin and W. Tan. Semi-supervised truth discovery. *In Proc. of the International Conference on World Wide Web (WWW'11),* pages 217-226, 2011.
26. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal Of The Royal Statistical Society, Series B*, 39(1):1–38, 1977**.**
27. J. Pasternack and D. Roth. Latent credibility analysis. *In Proc. of the International Conference on World Wide Web (WWW'13),* pages 1009-1020, 2013.

28. D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti, et al. Using humans as sensors: An estimation-theoretic perspective. *In Proc. of the International Conference on Information Processing in Sensor Networks (IPSN'14),* pages 35-46, 2014.

29. Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han. Resolving con flicts in heterogeneous data by truth discovery and source reliability estimation. *In Proc. of the ACM SIGMOD International Conference on Management of Data (SIGMOD'14),* pages 1187-1198, 2014

30. Amir Gandomi and Murtaza Haider "Beyond the hype: Big data Concepts, Methods and analytics", International Journal of Information Management (IJIM) ELSEVIER, 2015, pp: 137-144.

31. Li, Yaliang, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. "A Survey on Truth Discovery." arXiv preprint arXiv:1505.02463 (2015).

32. Fan Zhang, LiYu, Xiangrui Cai,Ying Zhang ,Haiwei Zhang, "Truth Finding from Multiple Data Sources by Source Confidence Estimation", 12th Web Information System and Application Conference, 2015.

AUTHOR PROFILE:

P. BASTIN THIYAGARAJ is working as an Assistant Professor in the Department of Information Technology, St.Joseph'scollege(Autonomous), Tiruchirappalli, TamilNadu, India. I am having 7 years of experience in teaching and 2 years in research.

Dr. A. ALOYSIUS is working as an Assistant Professor in the Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He has 16 years of experience in teaching and research. He has published many research articles in the National / International conferences and journals. He has acted as a chairperson for many national and international conferences. Currently, eight candidates are pursuing Doctor of Philosophy Programmed under his guidance.