ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Credit Card Fraud Detection Using Advanced Combination Heuristic and Bayes' Theorem

Sahil Hak[1], Suraj Singh[2], Varun Purohit[3]

Software Engineering Undergraduate, Department of Computer Engineering, Delhi Technological University, Delhi, India[1,2,3]

ABSTRACT: In today's increasingly electronic society and with the rapid advances of electronic commerce on the Internet, the use of credit card transactions have become the standard for Internet and Web-based e-commerce. In this paper, we present a credit card fraud detection model which takes into account present as well as the past behaviour. The detection model comprises of card validation via Luhn's algorithm, two initial probability assignments based on address mismatch and spending pattern, an advanced combination heuristic, spending-pattern database and Bayes' theorem. Advanced combination heuristic is an improvement that eliminates the conflict of existing Dempster-Shafer theory.

KEYWORDS: CCFDS; credit-card; Luhn's algorithm; Longest Common Subsequence; DBSCAN; advanced combination heuristic; Bayes' theorem

I. Introduction

With rapid advancement of e-commerce, use of credit cards for purchases has exponentially increased. Unfortunately, fraudulent use of credit cards has also become a source of crime. Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. Credit card fraud is also an adjunct to identity theft. According to the United States Federal Trade Commission, while identity theft had been holding steady for the last few years, it saw a 21 percent increase in 2008. However, credit card fraud, that crime which most people associate with ID theft, decreased as a percentage of all ID theft complaints for the sixth year in a row.[1]

According to a study conducted by ACI Worldwide and Aite Group released in June 2014, about 41 per cent of card users in India have experienced card fraud during 2012-2014. Study also highlighted that the highest rate of fraud on prepaid cards is experienced by consumers in India at 18 per cent.[2]

Financial institutions employ various fraud prevention models for tackling this problem. But fraudsters are adaptive, and given time, they devise several ways to intrude such protective models. Despite the best efforts of the financial institutions, law enforcement agencies and the government, credit card fraud continues to rise. Fraudsters nowadays may constitute of a very inventive, intellect and fast moving fraternity. Several techniques for the detection of credit card fraud have been proposed in the last few years. We briefly review some of them in Section II and propose our model in further sections.

II. Related work

Various existing fraud techniques majorly explore decision trees, genetic algorithms, clustering techniques and neural networks.

Recently, Syeda et al.[3] have suggested the use of parallel granular neural networks for speeding up the data mining and knowledge discovery process. Maes et al.[4] have outlined an automated credit card fraud detection system by Artificial Neural Network - ANN as well as Bayesian belief networks - BBN. They show that BBN gives better results related to fraud detection and the training period is faster whereas the actual detection process is substantially faster with ANN. The neural network based methods are, in general, fast but not so accurate an retraining the neural networks is quite taxing.

Chen et al.[5] propose a method in which an online questionnaire is used to collect questionnaire-responded transaction (QRT) data of users. Further it uses a support vector machine (SVM) trained with this data and the QRT models are used to predict new transactions. Chen et al.[6] have recently presented a personalized approach for credit card fraud detection that employs both SVM and ANN. It tries to prevent fraud for users even without any transaction data. However, these systems are not fully automated and depend on the user's expertise level.

Researchers have applied data mining for credit card fraud detection. Chan et al.[7] divide a large set of transactions into smaller subsets and then apply distributed data mining for building models of user behavior. The resultant base models are then combined to generate a meta-classifier for improving detection accuracy. Brause et al.[8] have explored the possibility of combining advanced data mining techniques and neural networks to obtain high fraud coverage along with a low false alarm rate. Use of data mining is also elaborated in the work by Chiu and Tsai[9]. They consider web services for data exchange among banks. A fraud pattern mining (FPM) algorithm has been developed for mining fraud association rules which give information regarding the features that exist in fraud transactions. Banks enhance their original fraud detection systems by using the new fraud patterns to prevent attacks. While data mining techniques are relatively accurate, they are inherently slow.

Aleskerov et al.[10] present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection[11]. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of wrong detections.

Credit card fraud detection by S Panigrahi et al.[12] suggests a fusion approach using Dempster-Shafer theory and Bayesian learning. Results from rule based filters are used and combined them using Dempster - Shafer adder. Later, these beliefs are strengthened by transaction history database and Bayesian learning.

Dempster has given a rule to combine evidences coming from different independent sources. However, Dempster's rule of combination has been criticized as sometimes it gives some illogical results. Therefore the above proposed model is susceptible to failures.

We use a better and more elastic combination rule by Tazid Ali et al.[13] discussed in further sections that eliminates the conflict problem of Dempster-Shafer theory.

III. Proposed Algorithm

The proposed model comprised of six steps. Firstly, Luhn's Test is used to validate card numbers. Then, two rules ie. Address Mismatch and Degree of Outlierness are used to analyze the deviation of each incoming transaction from the normal profile of cardholder. These two steps compute initial beliefs. The initial belief values are combined to obtain an overall belief by applying Advanced Combination Heuristic in step four. Step five looks into the spending history to extract characteristic information about genuine and fraud transactions. The overall belief is further strengthened or weakened in the final step using Bayes' Theorem, followed by recombination of the calculated probability with initial belief of fraud using advanced combination heuristic.

Step 1 : Luhn's Algorithm

Following standard algorithm is used to validate credit card numbers,[14]

1. Reverse the order of the digits in the number.
2. Take the first, third, ... and every other odd digit in the reversed digits and sum them to form the partial sum S1.
3. Taking the second, fourth and every other even digit in the reversed digits. Multiply each digit by two and sum the digits if the answer is greater than nine to form partial sums for the even digits.
4. Sum the partial sums of the even digits to form S2.

5. If S1+S2 ends in zero,then the original number is in the form of a valid credit card number as verified by the Luhn test.

For example, if the trial number is 49927398716,

1. Reverse the digits:
   61789372994
2. Sum the odd digits:
   6 + 7 + 9 + 7 + 9 + 4 = 42 = s1
3. The even digits:
   1, 8, 3, 2, 9
   Two times each even digit:
   2, 16, 6, 4, 18
   Sum the digits of each multiplication:
   2, 7, 6, 4, 9
4. Sum the last:
   2 + 7 + 6 + 4 + 9 = 28 = s2
5. S1+S2 = 70 which ends in zero which means that 49927398716 passes the Luhn's test.

Step 2 : Address Matching

The rule checks whether the Billing Address and Shipping Address match or not. This check does not guarantee whether a transaction is fraud or genuine. But if the two addresses match,the transaction can be classified as genuine with a high probability. Else,the transaction is labelled as suspect.

Suppose the address has the format: Road Number, House Number, Street Address, City Code, Pin Code. Let the billing address be: Road Number 76, House Number 24, Wall Street, New Jersey, 123214. If the billing house number is 76, and the user enters house number 73 as the shipping house number, the 2 numbers don't match. This check can be directly performed by comparing the equality of the 2 numbers.

But if the billing street address is 'Wall Street' and the shipping street address is 'Wll Strt',there is scope for matching, since the user might have entered a shorter form of the address. For this check to be successful, we have to initially store the billing address in it's longest form. For example, Billing Address should be stored as 'Wall Street' and not 'Wll Strt'.

1. Count the number of characters in Billing street Address and shipping street address. Label them as C1 and C2. If length of shipping street address is greater than 60% of the length of billing street address, we can proceed to step 2. Otherwise, we can assume that the addresses do not match.
2. Find whether the number of words of billing street address is equal to number of words in shipping street address. If equal, we can proceed to step 3. Else, the addresses do not match.
3. Find whether the first letter of every word in shipping street address is same as the first letter of corresponding word in billing street address. If yes, we can proceed to step 4. Else, addresses do not match.
4. Find the longest common subsequence of the two street address. If the longest common subsequence is same as shipping street address,addresses match. Else, they don't.[15]

Step 3 : Outlier Detection

A customer usually carries out similar types of transactions in terms of amount, which can be visualized as part of a cluster. Since a fraudster is likely to deviate from the customer’s profile, his transactions can be detected as exceptions to the cluster – a process known as outlier detection.

We have used DBSCAN[16] (Density Based Spatial Clustering of Application with Noise) to generate clusters, using transaction amount as attribute. Any incoming transaction amount,that does not belong to any cluster is detected as outlier. Such an observation gives evidence that the transaction could be fraudulent. We measure the extent of deviation of an incoming transaction by its degree of outlierness. If the average distance of the amount p of an outlier transaction T from the set of existing clusters is x, then its degree of outlierness d_outlier is given by,

d_outlier,

$1 - \frac{\mathcal{E}}{x}$ If Neighbouring points of P < MinPts

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

0 Otherwise

MinPts: Minimum number of points required in the epsilon-neighborhood of each point to form a cluster.
E(epsilon): Maximum radius of the neighborhood.

Step 4 : Advanced Combination Heuristic

More often, it is seen that available information is interpreted in probabilistic sense because probability theory is a very strong and well established mathematical tool to deal with objective uncertainty (i.e., uncertainty arises from heterogeneity or the random character of natural processes). However, it is clear that not all available information, data or model parameters are affected by objective uncertainty (i.e., nature of the data, information or parameters are random) and can be handled by traditional probability theory. Imprecision may occur due to scarce or incomplete information or data, measurement error or data obtain from expert judgment or subjective interpretation of available data or information. Thus, model parameters, data may be affected by subjective uncertainty. Traditional probability theory is inappropriate to represent subjective uncertainty (i.e., uncertainty arises from the partial character of our knowledge of the natural world). To overcome the limitation of probabilistic method, Dempster put forward a theory and now it is known as evidence theory or Dempster-Shefer theory (1976).

For the credit card fraud detection problem, Demspter-Shafer Theory is more relevant as compared to other fusion methods since it introduces a third alternative: ‘‘unknown”, along with the measure of confidence in each of the alternatives. It provides a rule for computing the confidence measures of three states of knowledge: fraud, fraud and suspicious (unknown) based on data from new as well as old evidence. Furthermore, in DST, evidence can be associated with multiple possible events unlike traditional probability theory where evidence is associated with only one event. As a result, evidence can be more meaningful at a higher level of abstraction. Hence, we use Dempster–Shafer theory for combining evidences for this problem. However, one of the shortcomings of DST is that, for evidences with a high degree of conflict, the modeling may not be accurate.

Given two mass functions m_1 and m_2, Dempster-Shafer theory also provides a combination rule for combining them, which is defined as follows,

$$m(h) = m_1(h) \oplus m_2(h) = \frac{\sum_{x \cap y = h} \; m1(x) * m2(y)}{1 - \sum_{x \cap y = \phi} \; m1(x) * m2(y)}$$

The above rule of combination produces unreasonable results as demonstrated in the following example, Deqiang et al. 2008 [17]. Zadeh had given a compelling example of erroneous result found by using D-S theory. Suppose two doctors examine a patient and agree that it suffers from either meningitis (M), contusion (C) or brain tumor (T). Thus frame of discernment = {M, C, T}. Assume that the doctors agree in their low expectation of a tumor, but disagree in likely cause and provide following diagnosis:

$$m_1(M) = 0.99,\ m_1(T) = 0.01$$
$$m_2(C) = 0.99,\ m_2(T) = 0.01$$

Based on Dempster rule of combination, we get the unexpected final conclusion of m(T) = 1.
It means that the patient suffers with certainty from brain tumor. This unexpected result arises from the fact that the two bodies of evidence (doctors) agree that the patient most likely does not suffer from tumor but are in almost full contradiction for the other causes of the disease.

Using the advanced combination heuristic[13],we can avoid the conflict of Dempster of Dempster Shafer theory.
Let (F, m_1) and (F, m_2) be two body of evidences given by two independent experts. Suppose A_1, A_2, ….., A_n be the focal elements. We will construct a combine body of evidence by assigning new bpa to each of the focal elements A_1, A_2, ….., A_n. For each A_i we associate a value m'(A_i) as,

$$m'(A_i) = m_1(A_i) + m_2(A_i) - m_1(A_i)*m_1(A_i)$$

This value can be interpreted as the combined evidence in support of the focal elements A_i . The motivation behind this formulation is the probability rule for union of two events. It also resembles that algebraic sum of fuzzy sets. Then the quantity m'(A_i) / {1 - m'(A_i)} can be interpreted as the odds ratio in favour of A_i, which can also be considered as evidence in support of A_i. Since we will deal with conflict situation in evidence theory the denominator of the above expression may be 0. So, we replace the denominator by 1 + {1 - m'(A_i)}, which gives,

$$m'(A_i) = \frac{1 - (1 - m1(Ai\ \ )) * (1 - m2(Ai\ \ ))}{1 + (1 - m1(Ai\ \ )) * (1 - m2(Ai\ \ ))}$$

Finally we assign bpa of A_1 as,

$$m(A_i) = \frac{m'(A_i)}{\sum_n m'(A_i)}$$

Using our proposed combined rule we solved the compelling example given by Zadeh in which the frame of discernment is Θ = {M,C, T} and corresponding bpas are $m_1(M) = 0.99$, $m_1(T) = 0.01$ and $m_2(C) = 0.99$, m_2 (T) = 0.01. Our method gives the bpa as m(M) = 0.494745, m(C) = 0.49745 and m(T) = 0.0051 which is more reliable than Dempster’s rule of combination.

Step 5 : Spending Pattern Database

It comprises of genuine Transaction Record(for individual customers from their past behaviour) and Fraud Transaction Record(from different types of past fraud data).We represent each history transaction by set of attributes containing information like card number, transaction amount and time since last purchase.

To capture the frequency of card use, we consider the time gap between successive transactions on the same card. The transaction gap is divided into Ten mutually exclusive and exhaustive events – D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}. Occurrence of each event depends on the time since last purchase (transaction gap) on any particular card. Let Y represents the time gap. The event D_1 is defined as the occurrence of a transaction on the same card within 15 hours of the last transaction which can be represented as,

D_1 = True | 0 < Y <=15
D_2 = True | 15 < Y <=30
D_3 = True | 30 < Y <=45
D_4 = True | 45 < Y <=60
D_5 = True | 60 < Y <=75
D_6 = True | 75 < Y <=90
D_7 = True | 90 < Y <=105
D_8 = True | 105< Y <=120
D_9 = True | 120< Y <=135
D_{10} = True | Y>135

We next compute $P(D_i|h)$ and $P(D_i|h')$ from the Fraud Transaction Record and the Good Transaction Record, respectively. $P(D_i|h)$ measures the probability of occurrence of D_i given that a transaction is originating from a fraudster and $P(D_i|h')$ measures the probability of occurrence of D_i given that it is genuine.

$P(D_i|h)$ = (Number of Occurrences of D_i in FTR) / (Number of transactions in FTR)
$P(D_i|h')$ = (Number of Occurrences of D_i on card k in GTR) / (Number of transactions on card k in GTR)
FTR : Fraud Transaction Record
GTR : Genuine Transaction Record
$P(D_i) = P(D_i|h) * P(h) + P(D_i|h') * P(h')$

Step 6 : Bayes’ Theorem

The idea of belief revision is that, whenever new information becomes available, it may require updating of prior beliefs. Bayes’ rule gives the mathematical formula for belief revision, which can be expressed as,

$$P(D_i|h) = \frac{P(Di\,|h) * P(h)}{P(Di)}$$

From Step 5,

$$P(D_i|h) = \frac{P(Di\,|h) * P(h)}{P(Di\,|h) * P(h) + P(Di\,|h') * P(h')}$$

Now, P(h) and $P(D_i|h)$ are combined using advanced combination heuristic discussed in step 4, to get the final belief of fraud. This final belief is now compared against the threshold values to determine whether the transaction is genuine, fraud or suspicious.

Methodology

The card is first validated by Luhn's algorithm is step 1. Incoming transaction is first handled by two rules, address matching and outlier detection. Probability values from these two rules are combined using advanced combination heuristics to get the initial belief of fraud P(h) for the transaction. Ω_L and Ω_U are the upper and lower threshold values for transaction classification into suspicious, genuine and fraud. These values can be learned over time or can be manually fed if sufficient information is available with the bank. If P(h) is less than Ω_L, the transaction is considered to be genuine and is approved. If P(h) is greater than Ω_U, the transaction is considered to be fraudulent and overhead safety precautions are taken. In case P(h) is in between Ω_L and Ω_U, card is labeled as suspicious.

In case the transaction is found to be suspicious, it is inserted in the suspect table. Credit Card Fraud Detection System determines which event D has occurred out of the ten Di's explained in the step 6 above, and retrieves the corresponding P(E|h) and P(E|h') values from the tables FTR (Fraud Transaction Record) and GTR (Good Transaction Record), respectively. The beliefs P(h|E) and P(h'|E) are next computed using Bayes' theorem. The initial belief of fraud is further strengthened by combining these probabilities with initial belief of fraud using advanced combination heuristic. Now accordingly, the transaction is treated upon using threshold valuesΩ_L, the lower threshold and Ω_U, the upper threshold. Fraudulent transactions are updated in Fraud Transaction Record and genuine transactions are updated in Good Transaction Record.

IV. SIMULATION AND RESULTS

Basic Probability Assignments,

1. Address Matching

If address mismatch occurs then there is a high probability that it is a fraud transaction and low probability that it is a genuine transaction. Alternatively, if the addresses match, then there is a high probability that it is a genuine transaction and low probability that it is a fraud transaction.

If address mismatch occurs,we assume the following basic probability assignments,

$m_1(h) = 0.6$

$m_1(h') = 0$

$m_1(U) = 0.4$

If $m_1(h)$ is set too high, the probability that a transaction is detected as fraudulent will go up. This improves the detection rate but also raises the number of false alarms. Similarly, if $m_1(U)$ is set high, the number of suspicious transactions goes up, increasing the number of misses. Thus, the values 0.6 and 0.4 have been chosen to maintain a balance.

If addresses match,we assume the following probability assignments,

$m_1(h) = 0$

$m_1(h') = 0.6$

$m_1(U) = 0.4$

2. Outlier Detection

For a transaction detected as an outlier, we make the following basic probability assignments using the degree of outlierness,

$m_2(h) = 1 - (E/x)$ (E , x hold meanings as discussed in step 3)

$m_2(h') = 0$

$m_2(U) = 1 - (1 - (E/x))$

The test run was conducted with the tailor made data which covered all the varied possibilities. Wide ranges of both the band pass filters of address mismatch and outlier detection were taken into consideration. Seven of the many sample runs of different categories of transactions were selected and are shown below[Table I]

S.No	Address Mismatch	Degree of Outlierness	Initial belief of fraud	Category of Transaction
1	No	0.86	0.736844	Fraud
2	Yes	0.77	0.440058	Suspicious
3	No	0.46	0.531902	Suspicious
4	No	0.12	0.356390	Suspicious
5	Yes	0.15	0.060305	Genuine
6	Yes	0.26	0.114321	Genuine
7	Yes	0.46	0.230716	Genuine

Table I : Test Run for various category of transactions. Address matching from Step 2. Degree of Outlierness from Step 3. Initial belief of fraud from advanced combination heuristics from Step 4. Category of transactions are fraud for score greater than 0.7, genuine for score less than 0.3 and suspicious if the score is in between.

As mentioned in part III of this paper, the algorithm needs to learn threshold values of fraud, suspicious and genuine transactions or these values can be fed directly, if are available with the bank. We assumed initial belief of fraud less than 0.3 to be genuine,and initial belief of fraud greater than 0.7 to be fraud.

In Table I, various cases are tested, both for address match and address mismatch (Step 2). Degree of outlierness is computed next using DBSCAN algorithm (Step 3). The probabilities from these two rules are combined using advanced combination heuristic (Step 4) to get the initial belief of fraud. In Case 1, initial belief of fraud is 0.736844,which is greater than 0.7, so the transaction is marked as fraudulent. In Case 5, 6 and 7, initial belief of fraud is 0.060305, 0.114321 and 0.230716 respectively. They all are less than 0.3, so the transactions are marked as genuine.

Scores of transaction 2,3 and 4 are 0.440058, 0.531903 and 0.356390 respectively. They all lie in the suspicious zone (between 0.3 and 0.7). Therefore, they undergo one more round where-in we find P(E/h) and P(E/h') using the spending pattern database and then compute the bayesian probabilities P(h/E) and P(h'/E). The bayesian probabilities are then combined with initial belief of fraud using advanced combination heuristic to get the final belief of fraud. This gives us the reevaluated category of transaction. We apply the same concept of thresholds here as well. This is shown in the table[Table II] below.

S.No	P(E/h)	P(E/h')	P(h/E)	P(h'/E)	Final belief of fraud	Reevaluated Category of Transaction
2	0.54	0.58	0.422	0.577	0.42530	Suspicious
3	0.28	0.61	0.341	0.658	0.43252	Suspicious
4	0.76	0.47	0.471	0.528	0.40925	Suspicious

Table II : Further round for suspicious transactions. P(E/h), P(E/h') from Step 5. P(h/E), P(h'/E) from Step 6. Final belief of fraud computed using advanced combination heuristic (Step 4).

REFERENCES

1. Federal Trade Commission, "Consumer Sentinel Network Data Book: January - December 2008", 26 February 2009.
2. ACI Worldwide, “Global Consumers : Losing Confidence in the Battle Against Fraud”, June 2014.
3. M. Syeda, Y.Q. Zhang, Y. Pan, “Parallel granular neural networks for fast credit card fraud detection”, Proceedings of the IEEE International Conference on Fuzzy Systems, 2002, pp. 572–577.
4. S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, “Credit card fraud detection using Bayesian and neural networks”, Proceedings of the First International NAISO Congress on Neuro Fuzzy Technologies, 2002.
5. R.C. Chen, M.L. Chiu, Y.L. Huang, L.T. Chen, “Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines”, Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, vol. 3177, October 2004, pp. 800–806.
6. R.C. Chen, S.T. Luo, X. Liang, V.C.S. Lee, “Personalized approach based on SVM and ANN for detecting credit card fraud”, Proceedings of the IEEE International Conference on Neural Networks and Brain, October 2005, pp.810–815.
7. P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo, “Distributed data mining in credit card fraud detection”, Proceedings of the IEEE Intelligent Systems, 1999, pp.67–74.
8. R. Brause, T. Langsdorf, M. Hepp, “Neural data mining for credit card fraud detection”, Proceedings of the International Conference on Tools with Artificial Intelligence, 1999, pp. 103–106.
9. C. Chiu, C. Tsai, “A web services-based collaborative scheme for credit card fraud detection”, Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004, pp. 177–181.
10. E. Aleskerov, B. Freisleben, and B. Rao, “CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection”, Computational Intelligence for Financial Eng., pp. 220-226, 1997.
11. M.J. Kim and T.S. Kim, “A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection,” Proc. Int’l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.
12. Suvasini Panigrahi, Amlan Kundu, Shamik Sural, A.K. Majumdar, “Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning” , Information Fusion 10 (2009) 354–363.
13. Tazid Ali, Palash Dutta, Hrishikesh Boruah, “A New Combination Rule for Conflict Problem of Dempster-Shafer Evidence Theory”, International Journal of Energy, Information and Communications, Vol. 3, Issue 1, February, 2012.
14. Hans P Luhn, ”Computer for Verifying Numbers”, US Patent 2,950,048, August 23, 1960 (Public Domain)
15. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein (2001). "15.4 - *Introduction to Algorithms* (2nd ed.)”, MIT Press and McGraw-Hil,. pp. 350–355, ISBN 0-262-53196-8.
16. Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M., eds.,”A density-based algorithm for discovering clusters in large spatial databases with noise”, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. ISBN 1-57735-004-9.CiteSeerX: 10.1.1.71.1980.
17. H. Deqiang, H. Chongzhao, and Y. Yi, “A modified evidence combination approach based on ambiguity measure”, Proceedings of the 11th International Conference on Information fusion, (2008), pp.1-6.