Correlation of complex evidences and link discovery

Boris Kovalerchuk
Dept. of Computer Science, Central Washington University
Ellensburg, WA 98926, USA, borisk@cwu.edu
Evgenii Vityaev
Institute of Mathematics, Russian Academy of Science,
Novosibirsk, Russia. 630090 vityaev@math.nsc.ru

Abstract

The classical statistical correlation is an efficient technique for linking simple numerical data sets via a single correlation coefficient R. Correlation of complex evidences that involves structured objects, text and data in a variety of discrete and continuous scales (nominal, order, absolute and so on) requires development of a new technique. Complexity of this task dictates the need to combine the statistical techniques with other mathematical techniques. This paper outlines design of such a new technique called Hybrid Evidence Correlation (HEC).

Often any individual evidence does not reveal a suspicious pattern and does not guide investigation in forensic accounting and other forensic fields. In contrast correlation of two or more evidences with each other and background knowledge can reveal a suspicious pattern. A new area of Link Discovery (LD) emerged recently is a promising new area for such tasks. Potential applications of link discovery range from basic science to a variety of practical forensic tasks. Currently LD mostly relies on deterministic graphical techniques. Bayesian probabilistic and causal networks are another relevant techniques. Both techniques need further development to handle rare events. This paper combines first-order logic (FOL) and probabilistic semantic inference (PSI) for designing HEC. The approach is illustrated with an example of discovery of suspicious patterns. Computational efficiency and completeness of the algorithm is justified by a computational experiment and a theorem.

Main concepts of the approach are: (1) a set of evidences D described by a set of attributes {A} and predicates {P} with two or more arguments; (2) a domain ontology in the form of a hieratical taxonomy of terms starting from {A} and {P} as terminal nodes; (3) formalized definitions of normal patterns, {N} and suspicious patterns, {S} in terms of FOL and PSI; (4) classification of patterns (statistically significant vs. insignificant ones to capture important rare events), (5) a generator of potential suspicious patterns (hypotheses generator), G, (6) an evaluator of hypotheses/patterns E and (7) a selector of suspicious patterns L.

The approach was successfully tested for detecting transactions fraud on synthesized data. Data contained 6 attributes of a transaction: seller, buyer, seller, buyer types, sold item, and amount, price and date. The data type is char (alphanumeric codes) for a seller, buyer, type of seller, type of buyer, and sold item.

1. Introduction

What is a forensic accounting? "A: By combining the definitions, from the Webster Dictionary, of the terms forensic medicine and accounting we can define forensic accounting as: An accounting method that deals with the relation and application of the system used to record and summarize business and financial transactions to a legal problem" [6].

Several types of financial forensic services are identified in [1] (see Table 1). Similar categories are identified in [4] as fields of forensic accounting: Fraud risk assessments; Background checks; Information security risk assessment, Asset tracing, End-use monitoring, Vendor monitoring, and Money laundering compliance program. Asset tracing and recovery include tracing and identifying client assets that are in the unlawful possession or control of third parties; gathering intelligence on the third parties; recovery of assets through civil remedies [4]

Table 1. Financial forensic services [1].

Special Investigations	**Dispute Advisory**	**Transaction Dispute Management**
Fraud investigation and advice A cash or purchasing fraud using digital images of computers, monitor stand-alone pc's or networks and analyze massive amounts of data.	*Damages evaluation* (loss of profits, economic loss and consequential loss): breach of contract and intellectual property rights, misrepresentation, profess-sional negligence,partnership disputes, contentious business valuations, shareholder disputes.	*Pre completion advice on accounting aspects of sale and purchase agreements* identification of risk areas in the pre-transaction stages of an agreement, advising on the accounting aspects of a sale and purchase agreement.
Asset tracing and recovery (tracing assets across borders and interviewing and investigating suspects in many jurisdictions.)	*Auditor and accountant negligence*	*Review of completion accounts:* develop arguments and assess the merits of proposed adjustments, assist in negotiations and advise on the merits and demerits of settlement compared to expert determination.
Special purpose investigations e.g., investigations of anomalies such as balance sheet black holes or "one off" situations that require investigation and clarification.	*Failure of commercial arrangements* including supplier / buyer disputes and commercial interpretation of contracts.	*Completion accounts disputes on* involved in a sale or purchase agreement.
Anti - money laundering and other regulatory investigations (reviewing and implementing the requirements of the Money Laundering Regulations 1993 and the Financial Services Authority Sourcebook and others).	*Insurance claims investigation and quantification:* claims arising from business interruptions, product and public liability, fidelity guarantee, and personal injury	*Expert determination: to* prepare submissions to the Independent Expert and to resolve the matters in dispute.
Forensic IT Services (incident response and computer analytics forensic data recovery, tracking intruders on computer networks, data mining, analysis and manipulation).		*Breach of warranty claims*: establishing the nature of a warranty breach, cause and affect this entails, evaluating and quantify the value of a breach, anticipating the possibility of litigation.

Traditionally focus of anti-money laundering investigations has been the reporting of large cash transactions. Now the focus in money laundering investigations is changing from illegal business and drug trafficking to terrorism: “Terror funding presents an even greater challenge to the financial system since it can comprise both laundered money and clean money” [2]. From financial viewpoint the illegal business and drug trafficking **make dirty money appear clean** and terrorism **makes clean money (**e.g., charity money) **dirty and uses laundered money.** Below Figure 1 from [2] illustrates this difference.

This figure operates with three stages of money laundering:

Placement – the physical disposal of the initial proceeds derived from illegal activity.
Layering – separating illicit proceeds from their source by creating complex layers of financial transactions designed to disguise the audit trail and provide anonymity.
Integration – the provision of apparent legitimacy to criminally derived wealth.

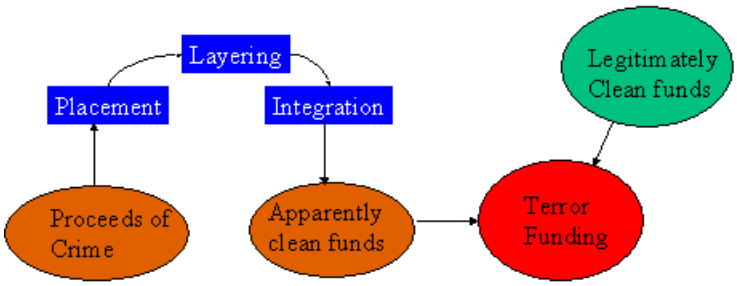

Figure 1. Modern money laundering diagram [2].

Modern money laundering investigations need to track [2]:

1. electronic funds (legitimate or otherwise);
2. criminal funds are already in the system and new cash placement;
3. funds placed in the less regulated countries and then wired into the better regulated jurisdictions;
4. the use of front companies that permits the criminal payments be made without using cash;
5. payment for a shipment of drugs disguised as the delivery of non existent goods or services; or payment for goods at a price in excess of the market price.

There are several challenges in **automated transaction monitoring systems** [2]:

- The identification of suspicious electronic transactions (it is considerably more difficult than the Identification of large cash transactions);
- Effectively highlight the unusual transactions using inexpensive, simple rules based systems and customer profiling;
- the reduction in the number of 'false positive' suspicious transactions by using profiling, statistical techniques, neural networks, fuzzy logic and genetic algorithms .
- Analysis of a larger picture (each institution may be seeing only a small part of a larger picture.[3]).

Current (since Sept. 11) practice: cross-checking customer and account lists against names that appear on law enforcement lists of suspected terrorists and money launderers, names that are transliterated and thus spelled in multiple way [3]. Types of suspicious (abnormal) transactions include:

- the proceeds of the transaction are channeled through several related and/or unrelated accounts before being moved offshore,
- a transaction that appears to have no identifiable business purposes,
- the account holder seems unconcerned about commissions and fees, despite their impact on the economics of the underlying transaction. [3].
- correspondent banking transactions (banks must ensure that they are not, directly or indirectly, providing correspondent services to offshore shell banks [3])
- transferor insolvent after transfer or was insolvent at the time of transfer,
- transfer for less than arm's length or less than reasonably equivalent value [5]
- wire transfers where they never occurred before [7].

To meet these challenges link-analysis software for forensic accountants, attorneys and fraud examiners such as NetMap and Analyst's Notebook and others [7-9] have been developed and under development now.

2. Example

Let us consider several types of transaction patterns:

- a normal pattern (NP) – a Manufacturer Buys a Precursor & Sells Result of manufacturing (MBPSR);
- a suspicious (abnormal) pattern (SP) – a Manufacturer Buys a Precursor & Sells the same Precursor (MBPSP);
- a suspicious pattern (SP) – a Trading Co. Bought a Precursor and Sold the same Precursor Cheaper (TBPSPC);
- a normal pattern (NP) -- a Conglomerate Buys a Precursor & Sells Result of manufacturing (CBPSR).

Below we discuss how these patterns can be discovered automatically from an ordinary or distributed transactions database using a statistical technique. We assume that DB contains transactions with attributes such as: seller, buyer, item sold, amount, cost and date (see illustration in Table 2).

Table 2.Transactions records

Record ID	seller	Buyer	Item sold	amount	cost	Date
1	Aaa	Ttt	Td	1t	$1000	03/05/99
2	Bbb	Ccc	Td	2t	$1000	04/06/98
3	Ttt	Qqq	Td	1t	$1000	05/05/99
4	Qqq	Ccc	Pd	1.5t	$1000	05/05/99
5	Ccc	Ddd	Td	2.0	$2000	08/18/98
6	Ddd	Ccc	Pd	3.0	$4000	09/18/98

Next information about types of companies and items sold is also partially available (see Tables 2 and 3).

Table 3. Company types and Item types

company name (seller/buyer)	Company type	Item	Item type in process PP
Aaa	Trading	Td	precursor
Bbb	Unknown	Pd	Product
Ccc	Trading	Rd	Precursor
Ttt	Manufacturing		
Ddd	Manufacturing		
Qqq	Conglomerate		

We need to assemble a new table (see Table 4) from tables 1-3 to reveal suspicious patterns in records. None of tables 2-3 indicate this individually. Table 4 also does not indicate suspicious patterns immediately. But we can map each pair of records in table 4 to patterns listed above using a pattern-matching algorithm *A* that analyzes pairs of records in table 4.

Table 4

Record ID	Seller	Seller type	Buyer	Buyer type	Item sold	Item type	Amount	Price	Date
1	aaa	trading	Ttt	manuf.	Td	Precursor	1t	$1000	03/05/99
2	bbb	unknown	Ccc	trading	Td	Precursor	2t	$2003	04/06/98
3	ttt	manuf.	Qqq	Congl.	td	Precursor	1t	$1000	05/05/99
4	qqq	Congl.	Ccc	trading	pd	Product	1.5	$2000	06/23/99
5	ccc	Trading	Ddd	Manuf.	td	Precursor	2.0	$2000	08/18/98
6	ddd	Manuf	Ccc	trading	pd	Product	3.0	$4000	09/18/98

Thus we can map pairs of records in table 4 into patterns:
A(#5,#6)=MBPSR, that is a manufacturer bought a precursor and sold product (normal pattern);
A(#1,#3)= MBPSP, that is a manufacturer bought a precursor and sold the same precursor (suspicious pattern);

A(#2,#5)= TBPSPC, that is a trading Co. bought a precursor and sold the same precursor cheaper (suspicious pattern).

Now let us assume that we have a database of 10^5 transactions as in table 1. Then table 4 will have all pairs of them, i.e., about $5*10^9$. Statistical computations can reveal a distribution of these pairs into patterns as shown in table 5.

Table 5

Pattern	Type	Frequency, %	Approximate number of cases
MBPSR	normal	55	$0.55*5*10^9$
MBPSP	suspicious	0.1	100
CBPSR	normal	44.7	$0.44*5*10^9$
TBPSPC	suspicious	0.2	200

Thus we have 300 suspicious transactions. This is 0.3% of total number of transactions and about $6*10^{-6}$% of total number of pairs analyzed. It shows that finding such transactions is like finding a needle in hay.

Finding all suspicious patterns illustrated in section 2.1 is a **computational challenge** for large and distributed databases, but an underlying algorithm *A* is relatively simple (see pseudo-code in table 6). This is because we have only two suspicious patterns/hypotheses descriptions defined in advance in terms of DB attributes.

Table 6. Algorithm for finding records that match suspicious patterns MBPSP and TBPSPC.

1. Form an SQL-query (Q1) to DB to retrieve pair of records that satisfy MBPSP
   1.1. Expand Table 1 with data from Tables 2 and 3 (make Table 4);
   1.2. make an SQL-query to find a pair (MBP record, and matching SP record) in Table 4.
2. Form an SQL-query (Q2) to DB to retrieve pair of records that satisfy TBPSPC;
   2.1. Use table Table 4 formed in 1.1;
   2.2. make an SQL-query to find a pair (TBP record, and matching SPC record) in Table 4.
3. Run query Q1 in a DB;
4. Run query Q2 in a DB.

The number of potential normal and abnormal types of patterns can be much larger and **automatic generation of patterns/hypotheses descriptions** is a major challenge that we are addressing in this paper. Thus our major question is: How to generate **automatically suspicious patterns/hypotheses using DB.** This includes generating MBPSP and TBPSPC descriptions automatically. Here we do not assume that we already know that MBPSP and TBPSPC are suspicious. One can ask: "Why do we need to discover these definitions (rules) automatically?" A manual way can work if the number of types of suspicious patterns is small and an expert is available. For multistage money laundering transactions it is difficult to accomplish manually. It is possible that many laundering transactions have been processed before money went to offshore.

Our **approach** to identify suspicious patterns is discovering **highly probable patterns** and **negating** them. We suppose that a highly probable pattern should be **normal**.

In more formal terms the **main statement/hypothesis (MH)** is:

If Q is a highly probable pattern (>0.9) then Q constitutes a normal pattern and not(Q) ***can*** *constitute a suspicious (abnormal) pattern.* (1)

Table 7 outlines an algorithm based on this hypothesis to find abnormal (suspicious) patterns. The algorithm is based first-order logic (FOL) and probabilistic semantic inference (PSI). More mathematical detail and a theorem on computational efficiency can be found in [11].

Table 7. HEC algorithm steps for finding suspicious patterns based on MH.

1	Discover patterns as Horn clauses, $A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n$; e.g., MBP $\Rightarrow$SR.
	Generate a set of predicates **P**=$\{P_1,P_2,\ldots,P_m\}$ and first order logic (FOL) sentences $A_1,A_2,\ldots,A_n$ based on **P**.
	Compute a probability that A1&A2&…&An-1$\Rightarrow$ An is true on a given database. This probability is computed as a conditional probability P(An/A1&A2&…&An-1)=N(An/A1&A2&…&An-1)/N(A1&A2&…&An-1,&An), where N(An/A1&A2&…&An-1) is the number of A1&A2&…&An-1& An cases and N(A1&A2&…&An-1) is the number of $A_1\&A_2\&\ldots\&A_{n-1}$ cases.
	Compare $P(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)$ with T=0.9. If $P(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)>T$ then a data base is 'normal". Threshold T can have another value too. For instance, P(MBP$\Rightarrow$ SR) can be 0.998
	Test statistical significance of $P(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)$. We use Fisher criterion [Kovalerchuk, Vityaev, 2000] to test statistical significance
2	If database is "normal" $(P(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)>T=0.9$ and $A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n$ is statistically significant then negate $A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n$ to produce $\neg(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)$.
3	Compute probability $P(\neg(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n))$ as $P(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow \neg A_n)=$ $= 1- P(\neg(A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow A_n)$. In the example above it is 1-0.998=0.002.
4	Analyze DB records that satisfy $A_1\&A_2\&\ldots\&A_{n-1}\ \&\ \neg A_n$. For instance, really suspicious records satisfy property $A_1\&A_2\&\ldots\&A_{n-}\&\neg A_n$, but other records also can satisfy this property too.

To minimize computations we generate randomly a **representative part** of all possible pairs of records such as shown in table 4. Then an algorithm finds highly probable (P>T) Horn clauses. Next these clauses are negated. After that a full search of records in DB is performed to find records that satisfy negated clauses. According to our hypothesis (1) this set of records will contain suspicious records and search of actually "red flag" transactions will be significantly narrowed.

Use of the property of **monotonicity** is another tool we use to minimize computations. The idea is based on a simple observation:

If $A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow$ B represents a suspicious pattern then $A_1\&A_2\&\ldots\&A_{n-1}\ A_n\Rightarrow$ B is suspicious too.

Thus one does not need to test clause $A_1\&A_2\&\ldots\&A_{n-1}\ A_n\Rightarrow$ B if $A_1\&A_2\&\ldots\&A_{n-1}\Rightarrow$ B is already satisfied. We used this approach successfully for medical applications [12,13].

In general terms main concepts of the approach are:

(1) a set of evidences D described by a set of attributes {A} and predicates {P} with two or more arguments;

(2) a domain ontology in the form of a hieratical taxonomy of terms starting from {A} and {P} as terminal nodes;

(3) formalized definitions of normal patterns, {N} and suspicious patterns, {S} in terms of FOL and PSI;

(4) classification of patterns (statistically significant vs. insignificant ones to capture important rare events),

(5) a generator of potential suspicious patterns (hypotheses generator), G, and

(6) an evaluator of hypotheses/patterns E and (7) a selector of suspicious patterns L.

4. Testing hypothesis

One of the technical aims of this paper is to design tests for this statement and test it in a simulation experiment. We designed two test experiments:

1) Test 1: Generate a relatively large table 4 that includes a few suspicious records MBPSP and TBPSPC. Run a statistical data mining algorithm (MMDR [11] to discover as many as possible highly probable patterns. Check that patterns MBPSR and CBPSR are discovered among them. Negate MBPSR and CBPSR to produce patterns MBPSP and TBPSPC. Run patterns MBPSP and TBPSPC to find all suspicious records consistent with them.

2) Test 2: Check that other found highly probable patterns are normal; check that their negations are suspicious patterns (or contain suspicious patterns).

A positive result of test 1 will confirm our hypothesis (statement) for MBPSR and CBPSR and their negations. Test 2 will confirm our statement for a wider set of and patterns. In this paper we report results of conducting test 1. The word *"can"* is the most important in our statement/hypothesis. If majority

of not(Q) patterns will be consistent with an informal and intuitive concept of suspicious pattern then this hypothesis will be valid. Otherwise if only few of not(Q) rules (patterns) will be intuitively suspicious than the hypothesis will not be of much use even it is formally valid.

A method for test 1 contains several steps:

- Create a Horn clause: MBP $\Rightarrow$SR.
- Compute a probability that MBP $\Rightarrow$ SR is true on a given database. Probability P(MBP =>SR) is computed as a conditional probability P(SR/MBP)=N(SR/MBP)/N(MBP), where N(SR/MBP) is the number of MBPSR cases and N(MBP) is the number of MBP cases.
- Compare P(MBP $\Rightarrow$ SR) with 0.9. If P(MBP $\Rightarrow$ SR)>0.9 then a data base is 'normal". For instance, P(SR/MBP) can be 0.998.
- Test statistical significance of P(MBP $\Rightarrow$SR). We use Fisher criterion [Kovalerchuk, Vityaev, 2000] to test statistical significance.
- If database is "normal" (P(MBP $\Rightarrow$ SR) >T=0.9) and P(MBP $\Rightarrow$SR) is statistically significant then negate MBP=>SR to produce $\neg$(MBP $\Rightarrow$SR). Threshold T can have another value too.
- Compute probability P($\neg$(MBP $\Rightarrow$SR)) as P(MBP $\Rightarrow$ $\neg$(SR)) = P($\neg$(SR)/MBP)=1- P(MBP$\Rightarrow$SR). In the example above it is 1-0.998=0.002.
- Analyze BD records that satisfy MBP and $\neg$(SR). For instance, really suspicious MBPSP records satisfy property MBP and $\neg$(SR), but other records also can satisfy this property too. For instance MBPBP records (a manufacturer bought a precursor twice) can be less suspicious than MBPSP.

Thus, if probability P(SR/MBP) is high (0.9892) and statistically significant then we can say that a normal pattern MBPSR is discovered. Then we suppose that suspicious cases are among cases where MBP is true but conclusion SR is not true. We can collect all such cases and start to analyze the actual content of the then-part of the clause MBP =>SR. We can discover that the set $\neg$SR contains variety of entities. Some of them can be very legitimate cases. Therefore, this approach does not guarantee that we find only suspicious cases, but the method narrows the search to a much smaller set of records. In the example above search is narrowed to 0.2% of the cases.

5. Experiment.

We generated two synthesized database with attributes shown in table 4. The first one does not have suspicious records MBPSP and TBPSPC. A second DB contains few such records. Using a Machine Method for Discovery Regularities (MMDR) [11] we were able to discovery MBPSR and CBPSR normal patterns in both DBs. MMDR method worked without any information in advance that these patterns are in data. See table 8.

Table 8. DB with suspicious cases.

Pattern	Probability P(A_1&A_2&…&$A_{n-1}\Rightarrow A_n$)	
	In DB without suspicious cases	In DB with suspicious cases
Normal pattern, MBP $\Rightarrow$ SR	> 0.95	>0. 9
Negated $\neg$(MBP $\Rightarrow$SR), MBP $\Rightarrow$ $\neg$(SR)	< 0.0.5	< 0.1
Normal pattern CBP=> SR	>0.95	> 0.9
Negated pattern $\neg$(CBP $\Rightarrow$ SR), CBP $\Rightarrow$ $\neg$(SR)	<0.05	< 0.05

In the database without suspicious cases negated patterns MBP $\Rightarrow$ $\neg$(SR) and CBP $\Rightarrow$ $\neg$(SR) contain cases that are not suspicious. For instance, MBP $\Rightarrow$BP, that is a manufacturer bought more precursors. The difference in probabilities for MBP $\Rightarrow$ $\neg$(SR) in two DBs points out actually suspicious cases.

In our computational experiments the total number of regularities found is 41. The number of triples of companies (i.e., pairs of transactions) captured by regularities is 1531 out of total 2772 triples generated in the experiment. Table 9 depicts the first 14 statistically significant regularities.

Table 9. Computational experiment: discovered regularities

#	Discovered regularity	Frequency
1	IF Seller_type = Manufacturing AND **Buyer__type** = Manufacturing THEN New_Item_type = product	72 / (6 + 72) = 0.923077
2	IF Seller_type = Manufacturing AND **New_Buyer__type** = Manufacturing THEN New_Item_type = product	72 / (6 + 72) = 0.923077
3	IF Seller_type = Manufacturing AND **Item_type** = precursor THEN New_Item_type = product	152 / (59 + 152) = 0.720379
4	IF Seller_type = Manufacturing AND **Item_type** = precursor AND 683.07 < **Price** THEN New_Item_type = product (with)	49 / (7 + 49) = 0.875000
5	IF Seller_type = Manufacturing AND Item_type = precursor AND **New_Price** < 47.25 THEN New_Item_type = product	10 / (0 + 10) = 1.000000
6	IF Seller_type = Manufacturing AND Item_type = precursor AND **Price_Compare** = 1 THEN New_Item_type = product	79 / (5 + 79) = 0.940476
7	IF Seller_type = Manufacturing AND 683.07 < **Price** THEN New_Item_type = product	120 / (55 + 120) = 0.68
8	IF Seller_type = Manufacturing AND 683.07 < Price AND **Buyer__type** = Trading THEN New_Item_type = product	32 / (5 + 32) = 0.864865
9	IF Seller_type = Manufacturing AND 683.07 < Price AND **New_Buyer__type** = Trading THEN New_Item_type = product	32 / (5 + 32) = 0.864865
10	IF Seller_type = Manufacturing AND 683.07 < Price AND **Item_type** = precursor THEN New_Item_type = product	49 / (7 + 49) = 0.875000)
11	IF Seller_type = Manufacturing AND **Price_Compare** = 1 THEN New_Item_type = product	182 / (92 + 182) = 0.664234
12	IF Seller_type = Manufacturing AND Price_Compare = 1 AND Buyer__type = Trading THEN New_Item_type = product	47 / (2 + 47) = **0.959184**
13	IF Seller_type = Manufacturing AND Price_Compare = 1 AND New_Buyer__type = Trading THEN New_Item_type = product	47 / (2 + 47) = **0.959184**
14	IF Seller_type = Manufacturing AND Price_Compare = 1 AND Item_type = precursor THEN New_Item_type = product	79 / (5 + 79) = **0.940476**

6. Conclusion

The Hybrid Evidence Correlation (HEC) is outlined in this paper. This technique advances statistical techniques to deal with complex (non-numeric) evidences that involve structured objects, text and data in a variety of discrete and continuous scales (nominal, order, absolute and so on).

Often any individual evidence does not reveal a suspicious pattern and does not guide investigation in forensic accounting and other forensic fields. In contrast correlation of two or more evidences with each other and background knowledge can reveal a suspicious pattern. A new area of Link Discovery (LD) emerged recently is a promising new area for such tasks. The paper shows potential application of HEC technique for forensic accounting. The technique combines first-order logic (FOL) and probabilistic semantic inference (PSI) for designing HEC. The approach is illustrated with an example of discovery of suspicious patterns in forensic accounting. Computational efficiency and completeness of the algorithm is justified by a computational experiment and a theorem.

7. References

1. Forensic Services: overview, 2002, Ernst and Young LLP, UK, http://www.ey.com/GLOBAL/gcr.nsf/UK/Forensic_Services_-_overview
2. Prentice, M.,Forensic Services - tracking terrorist networks,2002, Ernst & Young LLP, UK, http://www.ey.com/global/gcr.nsf/UK/Forensic_Services_-_tracking_terrorist_networks
3. Don Vangel and Al James Terrorist Financing: Cleaning Up a Dirty Business, the issue of Ernst & Young's financial services quarterly, Spring 2002. http://www.ey.com/GLOBAL/content.nsf/International/Issues_&_Perspectives_-_Library_-Terrorist_Financing_Cleaning_Up_a_Dirty_Business
4. Forensic Accounting, http://www.in.kpmg.com/services/services_assurance_nav3.html, 2002.
5. Rosenthal, H., Fraud and the Auditor In the Real World, Forensic Accounting Information and Education Center, LLC, 2001, http://www.askhal.com/fraud.html
6. IRS forensic accounting by TPI, 2002, http://www.tpirsrelief.com/forensic_accounting.htm
7. Chabrow, E. Tracking The Terrorists, Information week, Jan. 14, 2002, http://www.tpirsrelief.com/forensic_accounting.htm
8. How Forensic Accountants Support Fraud Litigation, 2002, http://www.fraudinformation.com/forensic_accountants.htm
9. i2 Applications - Fraud Investigation Techniques, http://www.i2.co.uk/applications/fraud.html
10. Evett, IW., Jackson, G. Lambert, JA , McCrossan, S. The impact of the principles of evidence interpretation on the structure and content of statements. Science & Justice 2000; 40: 233–239
11. Kovalerchuk, B., Vityaev, E., Data Mining in Finance: Advances in Relational and Hybrid Methods, Kluwer, 2000
12. Kovalerchuk, B., Vityaev E., Ruiz J.F., Consistent and Complete Data and "Expert" Mining in Medicine, In: Medical Data Mining and Knowledge Discovery, Springer, 2001, pp. 238-280.
13. Kovalerchuk, B., Vityaev E., Ruiz J.F., Consistent Knowledge Discovery in Medical Diagnosis, IEEE Engineering in Medicine and Biology Magazine, (Special issue on Data Mining and Knowledge Discovery), vol. 19, N, 4, July/August 2000, pp. 26-37.