Volume 2, Issue 1, January 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Representing Complex Counterfactual Sentences in Relational Databases

Toyin Enikuomehin
Department of Computer Science
Lagos State University,
Lagos, Nigeria

Abstract: **The need to aid computer representation of human activities cannot be overemphasized and this has led to attempts in capturing negative hypothesis for appropriate modelling of natural language representation in databases. Over time, many researchers have been carried out in presenting a human like computer essentially in the area of human computer interaction. This paper considers the complex nature of human sentences as embodied in the theory of counterfactuals and investigates a methodology for its representation such that a Natural Language interface can detect the structure of information contained in counterfactuals and execute such as appropriate, the paper extends the use of directed acyclic graph to analyze such structure. This paper presents a framework for further research into the understanding and presentation of counterfactuals against strict discreet domains.**

*Keywords; **Counterfactuals; causations; Natural Language; Databases***

I. INTRODUCTION

The task has always been seeking a man like computer; this expectation continues to grow exponentially as more people appreciate the capabilities of the computer itself. Several factors contribute to how well a computing process represents a humanist act; Prominent of this is the non crisp nature of the human thinking prospect against the strict discreet nature of the computer. The notion of counterfactuals is a topic in philosophy, studied and argued over by philosophers over the centuries. A formal, mathematical study of causation and counterfactuals has been initiated by the work of Pearl [1]. This has been a source of research in computer science and has led to development of several models. In this paper i propose a further consideration of the use of counterfactuals as a true representation of human behaviour over a crisp domain such as the relational database. Counterfactual thought refers to a mode of thinking that is literally contrary to fact. Counterfactual analysis has a long and distinguished history in comparative research. To some, counterfactual analysis is central to comparative inquiry because such research typically embraces only a handful of empirical cases[2].

In practice, counterfactual thought often supports speculation as to what might have been or what could have happened had some detail or event in the past occurred differently; as in the assertion,

"If I had been more careful, I might have won the price.

Counterfactuals are conditional statements in the subjective mood [3]. Formally they are presented as a grammatical form (which also relates to philosophy and logic). They involve conditions with false antecedents. In general computing, one of the starting points of the study of such reasoning is the observation that the conditional sentences of natural languages do not have a truth-conditional semantics. In traditional logic, the conditional "If A, then B" is true unless A is true and B is false. However, in ordinary discourse, counterfactual conditionals (conditionals whose antecedent is false) are not always considered true. Human reasoning is an extremely complex procedure which is able to handle such problems as counterfactual truth, which is difficult to explain in terms of logical theory. To achieve this, one needs to understand them in a more natural manner than just using the standard positivistic logic to do so. Counterfactual reasoning employs notions like the distance between possible worlds and the mental cost of reasoning. Other criteria for the truth of counterfactuals have been suggested, often within the framework of possible-worlds semantics. For example, the American philosopher David Lewis suggested that a counterfactual is true if and only if it is true in the possible world that is maximally similar to the actual one [4]. Intelligent systems require the ability to think and model the consequent of an action. As example, in game theory and its application, a game agent plans an optimization for each move in the gaming algorithm, thus game strategist needs to be able to reason that a potential move is likely to achieve a goal if it is executed, or otherwise. [5]

Cost functional is also included for such consideration as the strategist is expected to have the capabilities to forecast the cost of an alternative move which is the consequence. Such reasoning is a counterfactual reasoning and therefore a system capable of such reasoning can help in generating an idea strategy for gamers.

Counterfactuals have been shown to be very essential in the processes of Artificial Intelligence ranging from robotic control, gaming to semantic database querying. The semantics of a conditional $A > B$ are given by some function on the relative closeness of worlds where A is true and B is true, on the one hand, and worlds where A is true but B is not, on the other. Man has always expected the computer to perform almost all activities as being done by Man; this has led to exponential increase in the study of artificial intelligence.

Their logical analysis is problematic:

If NLC didn't go on strike, ASUU will

If NLC had not gone on strike, ASUU would have gone

Considering the two statement presented above, a plain conditional meaning could be derive from both statements if the other part of the phase is excluded or considered as an implication of the formal which means that both statement could be true. This generalization could not be derived using simple logical analysis. (Logic is generally accepted to be formal, in that it aims to analyze and represent the form (or logical form) of any valid argument type).

Some material tools used in representing truth functional are always introduced for easy of expression, such as '→' (or '⊃'), emplacing English presentation as If ..., then ...' If such gloss is correct when used, we must be ready to accommodate some other interesting features that might exist.

Consider the sentence given of the 'p → q' where p is said to imply q, this statement is always true when the antecedent, p, is false; and, secondly, it will always be true if the consequent, q, is true. But there are certainly some uses of 'If ..., then ...' which do not have these features. Let us consider the following statements:

If Obama had not won the ticket for democrats, then Gore would have won it. The antecedent of this sentence is false: obama got the ticket. But we still don't want to say that the sentence is true. If obama hadn't won the last election, Clinton would almost certainly have done so. There was virtually no chance of Gore winning. Thus the ascertion should not be implied in the sentence as 'If..., then ...' as a material conditional.

II. SOME EARLIER CONTRIBUTION

The exploration of counterfactuals dates back to time immemorial but this search was not formalized until the late 20th century when researchers began to inter-relate the concept of counterfactual to real living with computational features. In 1931, W. Churchill examined what would have happened if Robert lee won at the battle of Getty bury. Further presentation began to attract serious attention such as the Robert Fogel book in 1964 titled: railroad and American Economic growth: Essay in economic History, where he used quantitative methods to imagine what the U.S. economy would have been like in 1890 had there been no railroads [6]. In his hypothesis, he imagined that if railroad were never invented, then they would have been serious expansion on America large canal system and thus the pavements would have been improved. In the paper (*Plausible Worlds: Possibility and Understanding in History and the Social Sciences* (1991)by the Cambridge sociologist Geoffrey Hawthorn), the counterfactual history was carefully introduced into the academia. the publication contributed in the 1997 presentation of e *Virtual History: Alternatives and Counterfactuals* (1997), a collection of essays exploring different scenarios by a number of historians, edited by the historian Niall Ferguson. Ferguson has become a significant advocate of counterfactual history, using counterfactual scenarios to illustrate his objections to deterministic theories of history such as Marxism, and to put forward a case for the importance of contingency in history, theorizing that a few key changes could result in a significantly different modern world. Many briefs have been summitted to support the importance of counterfactuals. In [5], [6], it was pointed out that counterfactual reasoning is employed in many cognitive processes and counterfactual thinking requires simultaneously holding two different constructs of a given state of affairs [7].

Conflicts of understanding the main objective of counterfactual has long existed with some analyst insisting that counterfactual is much of which previous event is more relevant rather than what happened in the past.

III. COMPUTING COUNTERFACTUALS:

Conditional computing involves the use of protasis and apodosis. For protasis, we refer to the if clause and for apodosis, we refer to the then clause. An extension of the apodois is the else clause but not discussed in the domain of counterfactual. Given a counterfactual sentence, the protasis may or may not be true so therefore the apodosis may or may not be true. The apodosis is said by the speaker to be true if the protasis is true. A corresponding pair of examples with present time reference uses the present indicative in the *if* clause of the first sentence but the past subjunctive in the second sentence's *if* clause:

- *If it **is raining**, then he **is** inside.*
- *If it **were raining**, then he **would be** inside.*

in the first sentence the *if* clause may or may not be true; the *then* clause may or may not be true but certainly (according to the speaker) is true conditional on the *if* clause being true. Here both the *if* clause and the *then* clause are in the present indicative. In the second sentence, the *if* clause is not true, while the *then* clause may or may not be true but certainly would be true in the counterfactual circumstance of the *if* clause being true. In this sentence the *if* clause is in the past subjunctive form of the subjunctive mood, and the *then* clause is in the conditional mood.

Generally, it means that the presentation of the antecedent determine the truth nature of the consequent. This thought has been a benchmark in gaming application and its theoretical semantics where a condition of the form "If A, then B" is divided into two pairs, played with A and B, respectively. If a move A turns out to be true, it means that there exists a winning strategy in the game with A. The conditionality of B on A is thus implemented by assuming that this winning strategy is available to the verifier in the game with the consequent B. Counterfactuals play an essential role in practical reasoning.

Conceptual concept including metaphysical is an extension and explainable idea of counterfactual conditionals that is the epistemology of the former is a special case of the epistemology of the latter. In particular, the role of conceivability and inconceivability in assessing claims of possibility and impossibility can be explained as a special case of the pervasive role of the imagination in assessing counterfactual conditionals, an account of which is sketched. Thus scepticism about metaphysical modality entails a more far-reaching scepticism about counterfactuals. The account is used to question the significance of the distinction between a priori and a posterior knowledge [8]. Such knowledge is required to be captured in databases to preserve the relevance of the relational databases. The relational database is becoming increasingly less useful in a web 2.0 world due to lack of knowledge based features such as the ability to handle counterfactuals. The relational database model is great for storing information, but has not been properly extended to knowledge handling. By knowledge I mean information that has value beyond the narrow current conception of the given application. I mean information that can have enduring value form.

The reason the relational database doesn't represent knowledge very well is that the relational database is only good at storing objects and relationships between them when one fully understands exactly what objects and what relationships will be managed upfront. When you need to represent some new type of relationship between the objects in a relational database, it tends to fail, or be very difficult. In fact, the relational database isn't even particularly good at adding new types of objects to the database. Most relational databases actually have an upper limit on the types of objects, typically referred to as tables, which can be handled. Too many tables in a database schema are considered bad design.

The relational database is brittle but strong but does not handle knowledge appropriately however so much is being put in this direction.

IV. COUNTERFACTUALS, TRUTH TABLES AND SEMANTICS

Counterfactuals play an essential role in practical reasoning. Intelligent agents need to be able to reason counterfactually about the consequences of actions and events. The development of a formal semantics for counterfactuals by [4] and [8] stands as a major recent achievement in philosophical logic. I present again, the Lewis' counterfactuals thinking and its semantic generation using tables. The thought of using implication in truth table is considered appropriate as this could model the initial knowledge required in relational databases. This begins by firstly considering the Cartesian counterfactual presented below:

if $S = \sqrt{x^2 + y^2 + z^2}$

is the distance from a point P (x, y, z) to the origin. We assume that given Po = (1, 3, 1) is the world. We investigate whether
$y = 4 \succ s = \sqrt{19}$. (1)

Our Cartesian structure implies that x and z hold at some particular values 1, 1. Therefore we would have
$s = \sqrt{1+16+1} = \sqrt{18}$ (2)
and (1) is therefore an untrue counterfactual. However the counterfactual $y = 3 \succ s = \sqrt{11}$ is true. A change of theory, i.e. of co-ordinate systems, e.g. $x' = x+0.1y$, $y' = y$, $z' = z$, changes which counterfactual true. This means that the counterfactuals of truth tables are of great importance in knowledge transfer between men and machine. One limitation of the truth table approach is that it is designed for causal conditions are simple presence/absence dichotomies. Such systems are the discreet crisp representation of machine operation in 0's and 1's as against the actual human representation thinking.

Formal logic seeks to abstract general principles of reasoning which are in some way independent of context. For example, if we know that sentence A is true, then we know that the sentence "Either A or B is true" is also true. We don't need to know exactly what A (or B) actually say in order to know that this is a valid piece of reasoning, where "valid" means that if the initial assumption (premise) is true, then the conclusion follows logically. We can write this as follows:

TABLE I: Logical representation based on premise attribute

Premise /conclusion	Initial state 1	Initial State 11	Output 1	Output 1	Index
P1	A TRUE	-	-	-	-
C1	EITHER A(t)	EITHER B(t)	A TRUE	B TRUE	ONE OF
P2	A TRUE	AND B TRUE	-	-	-
C2	AND A(t)	AND B(t)	A TRUE	A TRUE	BOTH
P3	A TRUE	THEN BTRUE	-	-	-
C3	IF A(t)	THEN B(t)	A T/F	B T/F	A(t) ->B(t)

Logical conclusions like the ones presented above have several interpretation and analysis to several people, majorly its meaning and implication is largely subjected the individual interpretation, consider that around the beginning of the twentieth century there was a movement to construct *all of mathematics* from this sort of abstract logic. That this movement ultimately failed in an interesting way that does not change my point that formal logic can go well beyond the obvious.

People use different types of shorthand for various logically important words like "and", "or", "not" and "if", the following shall be used:

A = A is true.
~A= A is not true.
AvB= Either A or B is true (or both)
A&B= Both A and B are true.
A->B= If A is true then B is true.

One way of explaining the exact content of each of these statements is in a *truth table* such as the following:

A	**B**	**~A**	**AvB**	**A&B**	**A->B**
T	T	F	T	T	T
T	F	F	T	F	F
F	T	T	T	F	T
F	F	T	F	F	T

TABLE 2: A logical truth table

In the above table, the negation of each of the following is presented; the table shows the result of their negation, union, intersection and implication. In practise t, the type of logical analysis presented above is an extraction in the algebraic set theory. If they are called conditional statements and counterfactuals that occur as result of its use within such context is called counterfactual.

Counterfactual conditional statements are pretty well impossible to put into formal logic. They seem to have an inextricable contextual component which makes them impossible to describe in a context less, abstract way[9]. Whereas the logical "A->B" doesn't require us to know anything about what A or B actually say, the statement "If A was true, B would be true" really needs context, and some idea of what A and B are, before we can use it. This is because we are speaking of some other possible (or sometimes impossible) world where A is true, and we need context and the elusive 'common sense' to tell us how much of the current world to imagine as changed before we consider whether B would be true in such a world. Semantically deductive meaning for such cases has been a subject of study for sometimes now. We shall represent the semantic analysis as postulated by Lewis and use that to propose a data structure model.

The Lewis presentation is of the form of a relation where Lewis' proposed a relation based approach which considers the order of a world in respect to similarity. This similarity is the reference to the actual world in use. The notion of closeness $v \leq w$ where v —the world is used such that

v is at least as close to the actual world w as the world v given as

· w <w v for any v = w

V. THE DATA STRUCTURE OF COUNTERFACTUALS FOR DATABASES.

Representing counterfactuals as shown above can be best implemented by generating an appropriate model like formula for such cases, where our world conditions can be represented as

if $X_1, \ldots, X_n$, then …Y...",

by doing the above, we transform the original equation hypothetical we therefore formulate some other conditions to see if the result is the expected result for the consequent as expected form the truth table. An advantage of the above is that it provides a measure for counterfactual dependence between variables, one replaces the original equation for each variable X_i with a new equation stipulating its hypothetical value while keeping the other equations unchanged, then one computes the values for the remaining variables to see whether they make the consequent true. This process is the process of surgical intervention for variables that are dependent on counterfactuals.

A variable *Y counterfactually depends* on a variable *X* in a model if and only if it is actually the case that $X = x$ and $Y = y$ and there exist values $x' \neq x$ and $y' \neq y$ such that replacing the equation for *X* with $X = x'$ yields $Y = y'$.

Considering the analysis made in [11] and Byrne et al [12], it was observed that pre-emption can exist over a large counterfactual domain requiring similar treatment. Let us adapt the analysis used by [13], Hitchcock presents a useful regimentation of this reasoning. He defines *a route* between two variables *X* and *Z* in the set *V* to be an ordered sequence of variables $<X, Y_1,\ldots,Y_n, Z>$ such each variable in the sequence is in *V* and is a parent of its successor in the sequence. A variable *Y* is *intermediate* between *X* and *Z* if and only if it belongs to some route between *X* and *Z*. Then he introduces the new concept of an active causal route:

The route $<X, Y_1,\ldots, Y_n, Z>$ is *active* in the causal model $<V, E>$ if and only if *Z* depends counterfactually on *X* within the new system of equations *E'* constructed from *E* as follows: for all *Y* in *V*, if *Y* is intermediate between *X* and *Z* but does not belong to the route <*X*, Y_1,…, Y_n, *Z*>, then replace the equation for *Y* with a new equation that sets *Y* equal to its actual value in *E*. (If there are no intermediate variables that do not belong to this route, then *E'* is just *E*.)

This is then modelled as:

- *A* = 1 if the assassin pours poison into the king's coffee, 0 otherwise;
- *G* = 1 if the bodyguard responds by pouring antidote into the coffee, 0 otherwise;
- *S* = 1 if the king survives, 0 otherwise.

And also suppose that we employ these structural equations:

- *A* = 1;
- *G* = *A*;
- *S* = (*A* & *G*) v (~*A* & ~*G*).

The directed graph for this model is depicted in Figure 1.

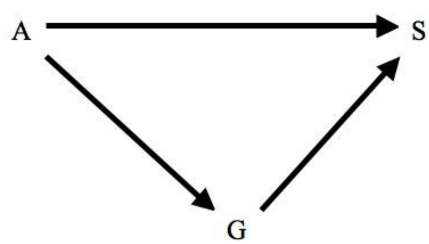

Figure 1: the directed graph representation of counterfactual consequence and antecedence

Now in theoretical computer science, a directed graph is a graph whose next point depends largely on the previous which is similar to the way at which relating antecedent to consequent in counterfactual works, we therefore investigate further that the consequent of any given counterfactual can be represented as a directewd graph of that condition. A directed graph cannot executive simultaneous operation thus only direct causation can be represented as the consequent in a given graph direction.

A directed (simple, unweighted) graph G is an ordered pair (V,E), where V is a set and E is a set of ordered pairs from V . That is, for any point(node or element X)

(∀x) ,(x ∈ V,E ↔ x ∈ V) → V = E

which is usually generalized as

E ⊆ (x,y): x,y ∈ V, x=?y

Directed graphs are suitable for modelling one-way streets, non-reflexive relations, hyperlinks in the World Wide Web, and so on. The notion of degree as defined not longer applicable to a directed graph. Instead, we speak of the in degree and the out degree of a vertex v in G, defined as |y{u 2 V : (u, v) 2 E}| and |{u 2 V : (v, u) 2 E}|, respectively. Finally, a graph can also be weighted, in the sense that numerical weights are associated with edges. Such weights are extremely useful for modelling distances in transportation networks, congestion in computer networks, etc. We will not dwell on weighted graphs in this course. The essence of this is that the study on counterfactuals will still be extended into its representation in database query. In such situations, queries in natural language involving counterfactuals can executed against relational databases. When such fit is achieved, the research for closing man machine relation has therefore been yielding some high level success. An acyclic directed graph such as a family "tree" has been known to have a representative format in relational databases. Since we have been able to model the counterfactual statement into a form of graph, the task is now to represent such graph into the relational database which will thus be fast in computing all ancestors of a node, and all descendants of a node. A way to do this is by initiating the following sql query:

```
CREATE TABLE nodes (
 id INTEGER PRIMARY KEY,
 name VARCHAR(10) NOT NULL,
 feat1 CHAR(1), -- e.g., age
 feat2 CHAR(1)  -- e.g., school attended or company
);

CREATE TABLE edges (
```

```
 a INTEGER NOT NULL REFERENCES nodes(id)
ON UPDATE CASCADE ON DELETE CASCADE,
 b INTEGER NOT NULL REFERENCES nodes(id)
ON UPDATE CASCADE ON DELETE CASCADE,
 PRIMARY KEY (a, b)
);

CREATE INDEX a_idx ON edges (a);
CREATE INDEX b_idx ON edges (b);
```

The big O for the number of select/insert/delete statements is also initiated to run query on ancestors and descendants, with ties broken by best big O for total runtime. This is to ensure a fast completion of the transformation process. One of the best way to represent the graph is by creating two index tables to handle the dimensions as shown above. With the above, we can now model a tree data structure into a relational database: parent-child model and same operation applies to a complex tree like a graph.

VI. CONCLUSION

This paper provides an outline the process of deriving some meaning and further implications from counterfactuals. We attempt to study the relevance of counterfactual in actual reasoning and further provide a tool that can assist in using counterfactuals in databases. The paper presents the use of directed graphs as a method for associating counterfactuals based on antecedent. The theoretical framework provided in the paper supports the proposal. Further work is required in this domain to implement the proposal as an executable query against relational databases. Achieving this will have great impact on user satisfaction in NL systems.

REFERENCE

1. J. Pearl, "Causation, action, and counterfactuals". Proc. TARK VI, pp. 51-73, 1996.
2. J. D. Fearon, "Counterfactuals and hypothesis Testing in Political Science". In: World Politics.43:169-195), 1991.
3. J. COLLINS *et al.*, "Counterfactuals,casuation and preemption", "*Causation and counterfactuals*". MIT Press, Cambridge, MA, 2004.
4. D. Lewis. "Causation". *Journal of Philosophy*, 70:556–567, 1973. Reprinted in [Lewis, 1986a, pp. 159–172].)
5. R.C. Stalnaker and R.H. Thomason, "A semantic analysis of conditional logic", Theoria 36,pp. 23-42, 1970
6. N. Goodman, "The problem of counterfactual conditionals", Journal of Philosophy, 44:113-128, 1947
7. J. McCarthy and P. Hayes, "Some philosophical problems from the standpoint of Artificial Intelligence". In Machine Intelligence 4, B. Meltzer and D. Michie (Eds.), Edinburgh University Press, Edinburgh, 1996.
8. J. Pearl, "Reasoning with cause and effect". Proc. IJCAI-99, pp. 1437-1449, 1999.
9. T. Costello and J. McCarthy, "Useful counterfactuals". *Electronic Transactions on Artificial Intelligence* Volume 3.3, 1999.
10. Christian Beyer, Alex Burri, " philosophical knowledge,its possibility and scope" (Unpublished)
11. M. L. Ginsberg, "Counterfactuals". Artificial Intelligence, 30, 35-79, 1986
12. R. M. J. Byrne, A. Tasso, "Deductive reasoning with factual, possible and counterfactual conditionals", Memory and Cognition, 27, 726-740. 1999.
13. C. Hitchcock, "The intransitivity of causation revealed in equations and graphs". *Journal of Philosophy*, 98:273–299, 2001.