A Method for Automating Token Causal Explanation and Discovery

Min Zheng and **Samantha Kleinberg**
Stevens Institute of Technology
Hoboken, NJ, 07030
{mzheng3, samantha.kleinberg}@stevens.edu

Abstract

Explaining why events occur is key to making decisions, assigning blame, and enacting policies. Despite the need, few methods can compute explanations in an automated way. Existing solutions start with a type-level model (e.g. factors affecting risk of disease), and use this to explain token-level events (e.g. cause of an individual's illness). This is limiting, since an individual's illness may be due to a previously unknown drug interaction. We propose a hybrid method for token explanation that uses known type-level models while also discovering potentially novel explanations. On simulated data with ground truth, the approach finds accurate explanations when observations match what is known, and correctly finds novel relationships when they do not. On real world data, our approach finds explanations consistent with intuition.

Introduction

While finding causal structures from data has been a core AI problem, finding why a particular event occurred (explanation) has received less attention. Causal explanation is what we do when finding the cause of a patient's seizure or assigning legal responsibility for a car crash. It is not enough to know what causes seizures in general, we must find the culprit for each specific event to provide effective treatment.

The general problem is where we have learned some causal relationships, and now have new observations we want to explain. A randomized trial may find that a drug causes headaches, and in post-market analysis we then aim to determine if it explains headaches and a few cases of insomnia. Current approaches try to link prior inferences to specific events, but cannot find explanations that are not in the model (e.g. the drug causing insomnia). More generally, this creates challenges when a variable is latent in the original data, the relationship is rare (and unlikely to be found at the type level), or the token event differs in its timing.

We propose a new method to identify novel explanations for events not fully explained by known causes. We build on and automate the method of (Kleinberg 2012), which enables explanation where token-level relationships need not exactly match the timing of type-level ones (e.g. finding aspirin relieves a headache in 29 minutes, even if known timing is 30-60 minutes) as it weights the difference between type and token. That approach gave a new measure of token significance, but did not show how to identify these weighting functions, was not evaluated experimentally, and could not discover new explanations. To handle when an event is not explained by known relationships, we build on the idea of defaults (Halpern and Hitchcock 2015) to discover token-level hypotheses while constraining the search space. We demonstrate on simulated data that our proposed method can explain cases that deviate from type-level knowledge (finding new explanations) and handle difficult cases such as causal chains and overdetermination, and on real data that it can discover intuitively correct explanations.

Related Work

One of the main explanation methods builds on the Bayesian network framework and is based on testing counterfactuals (Pearl 2000; Halpern and Pearl 2005). But if the user selects the wrong level of granularity for variables or the true relationship is not in the model, it cannot be inferred as a token cause (Halpern 2014). Needing user input means the approach cannot be fully automated, and explanations may be subjective. A model may encode that smoking causes lung cancer, but without timing details, a user must decide if smoking 2 months before developing cancer is a case of that relationship. The sequence of mechanisms method (Dash, Voortman, and Jongh 2013) links structural models and functional causal relationships to automate explanation by finding the likeliest path through a model, but the completeness of explanations is still governed by the model's.

Halpern (2008) modified the Halpern-Pearl framework for actual causation to incorporate normality (the most usual state for a variable) and rank token causes so that the most likely candidates are those that (counterfactually) involve switches from more atypical behavior to more typical behavior. If Jane usually waters a plant but then forgets, she is a much stronger cause of the plant's death than John, who never waters it. Halpern & Hitchcock (2015) distinguished between defaults (prior knowledge of what usually happens), typicality (related to knowledge and statistical frequency of events), and normality (which conveys value judgment). This can provide explanations that are more consistent with intuition, but the approach has not yet been automated due to the need for subjective assessments.

Within KR, understanding direct and indirect effects of

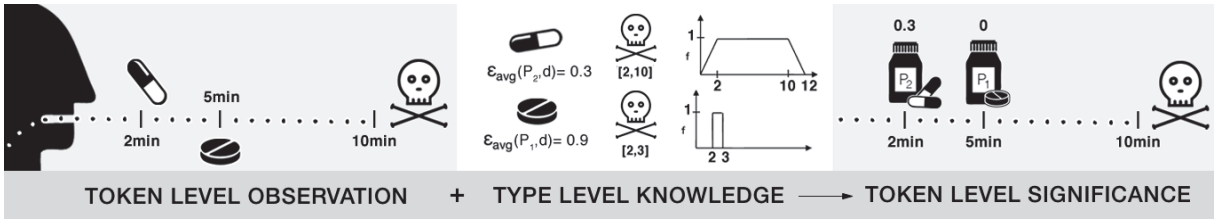


Figure 1: Explanation takes observations and type-level significance and ranks the significance of each token cause.

actions (McCarthy and Hayes 1969) is similar to prediction, but these methods project forward from an action, whereas we aim to look backward and find what led to it. Others can express the notion of "caused," (Giunchiglia et al. 2004) but involve reasoning in a causal logic rather than explaining events from data. Fault diagnosis, finding causes of unusual or incorrect system behaviors, is more related. Using logics such as the situation calculus and temporal extensions (Reiter 1996), these approaches can explain potentially complex events. Similarly, we build on probabilistic logic, but capture uncertainty rather than degrees of truth as in fuzzy logic for diagnosis (John and Innocent 2005). These approaches all rely on the completeness of the model, though, while we can discover explanations outside the set of known causes. That is, we could find a side effect is caused by a new drug, even if the effect was never observed in a clinical trial.

Background

A key limitation of existing methods is that token cases must exactly match type-level knowledge. Yet, some relationships may be unknown and the timing of a token case may differ. Kleinberg (2012) developed a framework to quantify the significance of token causes by weighting type-level significance by measurements that capture 1) uncertainty in token observations and 2) how much a token case differs from the type-level model. If they match exactly and our observations are certain, token-level significance is the same as the type-level significance of the causal relationship. When timings differ from what is known, significance is reduced.

While this approach can be used with other type-level causal significance measures, we use (Kleinberg 2012):

$$\varepsilon_{avg}(c_{r-s}, e) = \frac{\sum_{x \in X \setminus c} P(e|c \wedge x) - P(e|\neg c \wedge x)}{|X \setminus c|}, \quad (1)$$

with[1] $c \leadsto^{\geq r, \leq s}_{\geq p} e$, meaning that after c occurs, e occurs in time window $[r, s]$ with probability p. Causal significance, ε_{avg}, is the average difference a cause, c, makes to the probability of an effect, e, holding fixed each potential cause, $x \in X$, of e. Potential causes are $x : P(e|x) > P(e)$, and can be complex logical formulas. $P(e|c \wedge x)$ is the probability of e in time window $[r, s]$ after both c and x occur.

Definition of token significance With a type-level relationship $c \leadsto^{\geq r, \leq s}_{\geq p} e$, with associated significance $\varepsilon_{avg}(c_{r-s}, e)$, the significance of c at time t' as a token cause of e at time t relative to a sequence of observations $\mathcal{V}$ is:

$$S(c_{t'}, e_t) = \varepsilon_{avg}(c_{r-s}, e) \times P(c_{t'}|\mathcal{V}) \times f(c_{t'}, e_t, r, s). \quad (2)$$

An observation sequence is a set of timepoints with propositions true at each. The elements of the token causal significance (S) are the type-level causal significance (ε_{avg}), probability $P(c|\mathcal{V})$ of the cause given the observation sequence ($\mathcal{V}$), and function f that weighs how close the observed timings of c and e (t', t) match the type-level window $[r, s]$. When $t' \in [t - s, t - r]$, $f(c_{t'}, e_t) = 1$. Then f decreases monotonically outside that range and $f(c_{t'}, e_t) \in [0, 1]$. When the window is a strong constraint, f is a step function. For a certain observation ($P = 1$) where the timing matches the type-level ($f = 1$), the significance is that of the type-level cause. As observations become less certain, or differ more from the type level, significance decreases.

Fig. 1 shows a case where a person consumes two poisons and we want to know which caused their death. Using counterfactuals this case is overdetermined, as either poison can cause death, but we can handle it by accounting for time. At the type level we know p_1 causes death (d) in 2–3 time units with high significance. At other lags, it is completely ineffective, as shown by the steep drop in f on either side of the window. In contrast, p_2 is a weaker cause (lower type-level significance) but has a longer window of efficacy (2–10 time units), and its effect deteriorates less significantly outside this timing. Because the influence of p_1 drops off so steeply, we find its token-level significance, $\varepsilon_{avg} \times f(5, 10, 2, 3)$, is 0. Note that $P(p_1|\mathcal{V})$ is one since the observations are certain. Since p_2 is consistent with its known time window, its significance is the type-level significance, 0.3 (i.e. $\varepsilon_{avg}(p_2, e) \times P(p_2|\mathcal{V}) \times f(2, 10, 2, 10) = 0.3 \times 1 \times 1$).

The key challenges are that this approach needs information on how to weight the timing differences, and if a type-level relationship is unknown, it cannot be a token cause.

Method

We propose a new method building on (Kleinberg 2012) to find the significance of token causes while allowing some differences between observations and knowledge. First, we show how functions to weight similarity of events to type-level relationships can be learned during causal inference. Second, we build on ideas from (Mill 1843) to factor out known causes and (Halpern and Hitchcock 2015) to use changes from defaults to find causes for unexplained events.

Assumptions Our key assumption is that token causes are not latent: while a type-level causal relationship may be unknown, the cause itself is observed at the token level. So we

[1] This is a PCTL (Hansson and Jonsson. 1994) leads-to formula.

Algorithm 1 compute-f-function(V, T, H, l_{max}, D)

Input:
V, a set of variables; and T, length of the time series
H, a set of all causal relationships between $v \in V$
l_{max}, the maximum time lag to test
D, a $V \times T$ matrix indicating value of each variable
Output:
F, a list with a function f for each relationship in H

for each causal relationship $h \in H : c \rightsquigarrow^{\geq r, \leq s}_{\geq p} e$ **do**
  Calculate ε_{avg} for each lag $l \in [1, l_{max}]$ using D
  Normalize ε_{avg} by dividing by its maximum value
  Delete outliers where $\varepsilon_{avg} < 0$
  $f(c_{t'}, e_t, r, s) = 1$ when $t - t' \in [r, s]$
  $f(c_{t'}, e_t, r, s)$ for $t - t' \in [1, r) \bigcup (s, l_{max}]$ is fit to ε_{avg} with nonlinear least squares. Then add f to F.
return F

may not know c causes e, but have both measured for the case to be explained. Causal relationships may be complex (e.g. conjunction of variables), as allowed by the logic used, though here we search only for token causes composed of single variables to reduce complexity.

Finding weighting function from data

In eq. 2, we weight how closely observed timings match those identified at the type level, but we need to first determine such functions. Prior work suggested these can be created with background knowledge, but we propose they can be learned in a data-driven way using properties of the causal significance measure, ε_{avg}, and a smoothing procedure to maintain monotonicity of the function outside the type-level time window, and ensure it is in [0,1] (see fig. 1). Experimental results show this recovers the true underlying functions. The function f is similar to a membership function, though operating over time and capturing probability.

The process is shown in algorithm 1. For a given causal relationship, $c \rightsquigarrow^{\geq r, \leq s} e$, we use eq. 1 to calculate ε_{avg} individually for each lag $l \in [1, l_{max}]$, where l_{max} is a user set parameter (the maximum tested lag). One disadvantage of testing lags individually is that when data are not regularly sampled, we cannot guarantee that this will recover the correct time windows and that significance will increase and decrease monotonically before and after them. Consider data from visits to primary care physicians. Even if an actual effect is that a medication causes side effects in exactly 6 months, only a small set of observations will be at that exact time, and there may be none at all for some lags. Thus we do not use raw values for ε_{avg} since there may not be enough data to calculate them for every lag in $[r, s]$ and there may be small fluctuations around the true values. To identify the weighting function we normalize the ε_{avg} values so that f of those in $[r, s]$ is one. After removing outliers (negative ϵ_{avg}), we use nonlinear least squares to fit a curve to the values for ε_{avg}. This ensures that the range will be in [0,1], that the values will decrease monotonically on either side of $[r, s]$, and that the function is continuous.

Algorithm 2 find-novel-causes$(V, T, D, k, m, l_{max}, E)$

Input:
V, a set of variables; and T, length of the time series
D, a $V \times T$ matrix indicating value of each variable
k, number of explanations to select (for top-k)
m, max time lag for detecting state change
l_{max}, max time lag for finding novel explanation
E, set of explained events (pairs of form: $c_{t'}, e_t$)
Output:
E', token causal explanations for each event

for each $e \in V$ **do**
  Use E to remove explained events to get matrix D_e
  for each $c \in V$ **do**
    for each $l \in [1, l_{max}]$ **do**
      Get $P(e|c_l)$. When c has a default state, use instances of c where c is not in its default state or changed state up to m time units before.
for each non-null $D_e[e, i]$, $1 \leq i \leq T$ **do**
  $L_i \leftarrow [\,]$
  for each $D_e[c, j]$, where $i - l_{max} \leq j \leq i - 1$ **do**
    Add $(c_j, e_i, P(e|c_{j-i}))$ to L_i
  Add top k explanations for event e_i to E' (tuples: c_j, e_i, $P(e|c)$)
return E'

Discovering novel token causes

Prior approaches are limited by their reliance on an existing model. Yet models cannot capture all causes (particularly rare ones), and we may not have sufficient data to infer these from token cases. For instance, we do not want to wait until many people have a severe new side effect after a drug is on the market to identify that the drug is causing the side effect. We overcome this limitation by building on an observation similar to Mill's (1843) method of residues, and using concepts of normality to narrow down the set of possible causes.

When multiple causes and effects occur, Mill's method removes the known causes, then identifies which remaining factors explain the remaining effects. Thus, if acidic foods cause an upset stomach and we want to know what causes heartburn, we can essentially subtract the instances of acidic food and upset stomach and then examine only similarities between cases with heartburn, and differences between those with heartburn and those without. Our approach uses a similar idea of residues. We 1) remove events (i.e. variable e at time t) from the set to be explained if known causes can account for them (exceeding a significance threshold), then 2) test whether any variables can explain these remaining events. After step 1, we have a residual of what events cannot be explained by the type-level causes, which are then the set whose causes we aim to find in step 2. This only keeps track of which events require explanations – no variables are removed as potential causes of others.

First we must determine which variables can explain these events. Taking inspiration from graded causation as introduced by Halpern & Hitchcock (2015) to decide which variables are considered as novel explanations (token cause candidates), we observe that when an event is not explained by

what we know, the possible causes should be variables that are either not in their default state or those that have recently changed state. Thus even if every event of a headache occurs when an individual has normal body temperature, we can still avoid identifying temperature as a possible cause of headaches, while finding that prolonged increased heart rate from running can lead to hypoglycemia in people with diabetes. On the other hand, a change from cold rainy weather to average temperatures (the default for weather) may lead to an increase in outdoor activities due precisely to the improvement in weather. To find candidate token causes for each event to be explained, we first check whether a variable has a default state (e.g. body temperature is usually 37°C). Then, we identify which of these variables are not in that default state (e.g. most frequently observed state) during the tested time window before the event, or which have changed from non-default to default state within a window $[1, m]$ before the event (m is a user set parameter that can be inferred from the characteristics of the input data in Algorithm 2).

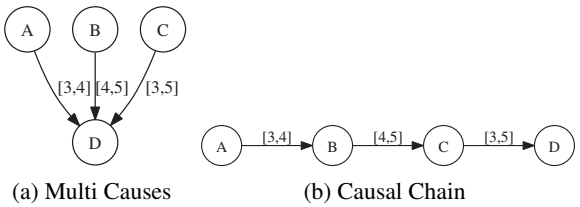
Explanation then proceeds independently for each effect. We remove events that are already explained, then look for commonalities in those that remain, using the conditional probability for discrete events, or conditional expectation for continuous-valued ones. That is, for unexplained cases of an effect e, we calculate $P(e|c)$ (or $E[e|c]$) for each candidate cause c in the time series, using a frequency-based approach. This value is calculated for each of a set of time lags, $l \in [1, l_{max}]$. The output is a ranked list of potential explanations for otherwise unexplained events. These cannot be guaranteed to be the true causes of each (as there may be a latent common cause of c and e). However, our assumption is that unknown relationships are primarily infrequent or weak, so stronger and more likely causes are known. Thus if $c \leftarrow d \rightarrow e$, either $d \rightarrow e$ will be known, or d will also be identified in this step and will have higher significance for e than c will. By restricting the candidates of novel explanations using defaults, we can avoid finding factors that are frequent but not causal (e.g. body temperature is normal).

Combined explanation process

The process, shown has two parts: 1) using type-level relationships to explain observed events, then 2) identifying novel token causes (or hypotheses for explanations) for unexplained cases (algorithm 2).

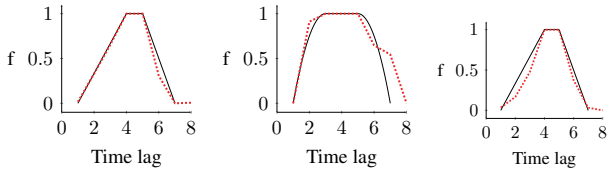
Step one: Explanation with type-level knowledge Using the approach of (Kleinberg 2012), we calculate the significance of all token-level explanations (eq. 2) then infer the weighting functions for each type-level relationship. For each effect to be explained, we evaluate the significance of each of its type-level causes for each specific instance. For each occurrence, using either a threshold for the significance value or keeping only the top k relationships, we then have a set of events for which no (sufficiently significant) causes have been found. This is the input to step two, which is shown in algorithm 2.

Step two: Identify novel causes For the remaining unexplained events, we rank explanations using the conditional probability of each such event given each observed variable, at each time lag (up to parameter l_{max}). Note that the set of events in this step is much smaller than that in step one. It is critical to remove explained events to avoid both confounding and failing to find a relationship due to the small number of occurrences being dwarfed by more reliable causes. Moreover, token cause candidates are only those that do not have a default state or in time window $[1, m]$ before the event are either not in their default state or have changed state.



(a) Multi Causes (b) Causal Chain

Figure 2: Causal structures for simulated data.



(a) RMSD=0.070 (b) RMSD=0.111 (c) RMSD=0.042

Figure 3: The true (solid) and inferred (dashed) weighting function (f) for causal relationships.

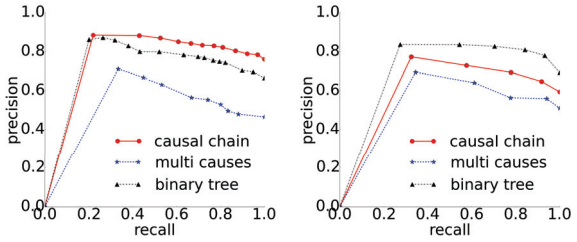
Experiments

We evaluate three areas on simulated data: finding weighting functions, explanation from a causal model, and finding new causes. Simulations yield ground truth for each event and whether new causes are genuine. We also apply the method to a real world bike sharing dataset. Other methods mainly rely on conceptual evaluation and do not have implementations, so quantitative comparisons are not feasible.

Simulated Data Generation

We create causal relationships (cause, effect, time window) and associated weighting functions f) and use these to generate observation sequences. Structures generated are shown in fig. 2 (plus a binary tree), and selected weighting functions in fig. 3. A variable with no causes may occur spontaneously, otherwise events occur if caused to. This happens at a time lag weighted by f, so lags in $[r, s]$ are most likely.

Parameters: Causal relationships have probability in $\{0.5, 0.65, 0.8, 0.95\}$. Each dataset has 5000 time points, and contains the causal structures plus 5 noise variables with no causes (2 for the binary tree), which have a probability of 0.01. The probability of the root cause in each structure occurring is: 0.15 (multi cause) and 0.25 (causal chain, binary tree). For experiments testing discovery of novel causes, we omit relationships from the type-level knowledge (dotted arrows in fig. 5) and give these relationships probability 0.2.



(a) Type=token (b) Type!=token

Figure 4: Precision-recall when type and token are consistent (left) and for novel token causal discovery (right).

Evaluation of weighting function

First, we evaluate how well the inferred weighting functions match ground truth, using root mean square difference (RMSD). Fig. 3 shows strong agreement. The only exception is one step function that goes immediately to zero outside the known window. Because of the granularity of the time measures, even though the values inferred are primarily one and zero, they are still connected by a gently sloping line. The mean RMSD is 0.122 (std. dev. 0.085).

Evaluation of inference where type = token

We first test if we can identify explanations when type and token are consistent. If we know A causes B in general, can we find the specific instances of A that explain each B?

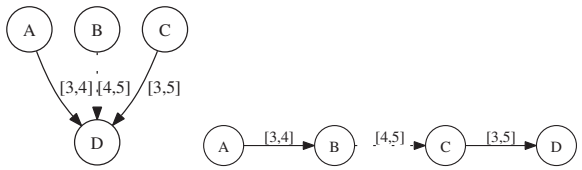
Fig. 4a shows precision-recall curves for each of the three structures generated by repeatedly lowering the threshold used for calling a relationship significant. We also use the rank ratio to determine how often the true cause of a specific event is in the top k most significant explanations. This is:

$$R = \frac{\sum_v \sum_{t \in T} |causes| / \min\{n, k\}}{T} \quad (3)$$

where $|causes|$ is the number of true causes in the top k, n the number of genuine causes for each instance of an effect, and T is the total number of occurrences of each variable v.

Multiple causes of a single effect (fig. 2a) In Table 2, the true cause is most significant >70% of the time, and this increases to 96.6% with $k = 3$. Our evaluation is strict in cases of overdetermination. In fig 2a, say A and B both occur, and D happens after. With our data generation process we can know that A actually caused D, so finding B causing that instance of D is a false positive. While we penalize the algorithm for finding B, this result is still better than counterfactual methods, where neither A nor B would be found. Similarly, when two variables contribute to an effect, even if we do not search over conjunctions we will find each individually (e.g. A and B each as causes of D).

We evaluated causal relationships with varying probabilities (0.95, 0.8, 0.65 and 0.5) to test the impact of causal strength. The difference between the highest and lowest AUC is only 0.07, suggesting this has minimal impact on accuracy. All following results are for 0.8.



(a) Multi Causes (b) Causal Chain

Figure 5: Causal structures with dashed edges indicating missing relationships, omitted from algorithm input.

Causal chain (fig 2b) This case has higher precision than for multiple causes (0.760 with recall of 1), and with $k = 1$, the rank ratio is 0.881. As each variable has one cause, there is one type of error: finding a different timing of the true cause as the explanation. Say B causes C and we observe B_2, B_3, C_4. If B at time 3 is the true cause of C at time 4, and we only find B_2 as the cause, that is a false discovery.

Binary tree The binary tree has three levels. This is a more difficult case with lower precision than the causal chain (0.664 with recall of 1). When $k = 1$, rank ratio is 0.839. The errors are similar to those of the causal chain case. In many cases due to the time windows of the relationships multiple instances of the variable occurring could each be responsible for the effect, but only one truly caused it, leading to an overdetermined case. Once again this is a case that cannot be handled by standard counterfactual methods.

Evaluation of discovery of novel token causes

Finally, we omit some relationships from the input: if the algorithm only knows A and C cause D (fig 5a), can it correctly infer that B causes some instances of D?

Multiple causes of a single effect Fig. 5a shows the updated model, where a dotted arrow indicates the relationship we aim to discover at the token level. This case has lower precision than the binary tree and causal chain cases (0.510 when recall is 1), because there are more potential causes for each effect, and when $k < 3$, it is possible that some known causes were not factored out. More variables are also candidates as explanations (e.g. noise variables).

Rank ratio is now more complex, because which cases are "unexplained" depends on which type-level causes we accept as significant. Table 1 has two values for k: columns are k for type-level relationships (i.e. value used to decide what events are explained), and rows are k for our token causal discovery. Accuracy goes up as type-level k does because with 3 causes of the actual effect, if k is small, the genuine cause may have lower significance than the other two in an overdetermined scenario. When $k = 3$, rank ratio is 98.8%.

Causal chain (fig. 5b) Precision is higher than the multiple cause case (0.594 when recall is 1), due to fewer confounders and less overdetermination. Tbl. 1 shows we find over 95% of the true causes when $k = 3$. False discoveries are mainly finding earlier links in the chain with longer lags.

Binary tree The updated model omits a relationship at each of the levels. This is a challenging case as one may find indirect elements of a chain with long time lags, instead of the true relationship and actual lag. Precision is higher than

Table 1: Result for novel explanation discovery. $k-$discover is the threshold for discovered causes, and k the rank threshold for type level relationships to be accepted as explanations.

	Multi Cause			Causal Chain			Binary Tree		
$k-$discover	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
1	0.701	0.737	0.749	0.688	0.688	0.688	0.803	0.819	0.825
2	0.936	0.946	0.947	0.870	0.870	0.870	0.948	0.957	0.960
3	0.988	0.988	0.988	0.952	0.952	0.952	0.970	0.978	0.979

Table 2: Rank Ratio when type and token are consistent.

top-k	Multi Cause	Causal Chain	Binary Tree
1	0.704	0.881	0.839
2	0.887	0.987	0.977
3	0.966	0.999	0.998

the other cases (0.692 with recall 1), possibly due to fewer noise variables (2 vs. 5). Rank ratio is 97.9% with $k = 3$.

Real world dataset

As a second evaluation, we apply our approach to a real-world bike sharing data set from the UCI data repository (Fanaee-T and Gama 2013). The data includes two years of bike sharing logs including environmental variables (e.g. temperature, humidity, wind), dates (e.g. holiday), and bike rental counts. We aim to discover explanations for bike rentals. We discretized continuous variables using three bins of equal width, then inferred type-level relationships, finding bad weather causes low rental count the same day as does low temperature. There are 194 high rental count events and 167 low rental ones. The type-level causes explained 19 (bad weather) and 71 (low temp) instances of low counts.

Next we aim to discover causes for instances not explained by the type-level causes (194 instances of high rentals, 77 of low). For weather, feels-like temperature, humidity, and wind speed, we set the default state as the average (temperature) or mild (wind, weather, humidity) states. The time window for checking whether a variable changes state is $m = [1, 4]$ days. We select only the top scoring cause for each instance. Causes of high rentals include: high temperature (121 cases), good weather (40), low wind speed (24), moderate humidity (7), and moderate temperature (2). For low rentals we find the causes: moderate temperature (49), high humidity (21), moderate humidity (6), and low humidity (1). Consistent with what one would expect, we found 95% of increases in rentals are due to favorable weather, while 67% of cases of decreases in rentals are due to poor weather conditions (e.g. low temperature, humidity, bad weather). Note that relying only on the type-level model only let us explain a small portion of the overall set of cases, while our approach for discovering explanations lets us identify explanatory hypotheses for all instances.

Conclusion

Our new automated causal explanation method can accurately 1) infer how to weight the timing of type-level relationships when evaluating token cases, 2) calculate significance of token causes, and 3) discover novel explanations for events that are not explained by type-level models. A key limitation of prior methods is the inability to explain events that are not instances of type-level relationships. Our approach leverages prior information to pare down the set of events requiring explanation: discovering new causes for events that are not consistent with the model, using the idea of residues and defaults. While causes cannot be guaranteed to be genuine in all cases, this approach can provide potentially unexpected hypotheses for further evaluation, and be extended to deal with latent token causes in the future.

Acknowledgments This work was supported in part by NSF Award #1347119.

References

Dash, D.; Voortman, M.; and Jongh, M. D. 2013. Sequences of Mechanisms for Causal Reasoning in Artificial Intelligence. In *IJCAI*.

Fanaee-T, H., and Gama, J. 2013. Event Labeling Combining Ensemble Detectors and Background Knowledge. *J. Prog Artif Intell.* 2:1–15.

Giunchiglia, E.; Lee, J.; Lifschitz, V.; McCain, N.; and Turner, H. 2004. Nonmonotonic Causal Theories. *Artificial Intelligence* 153(1):49–104.

Halpern, J. Y., and Hitchcock, C. 2015. Graded Causation and Defaults. *Brit. J. Philos. Sci.* 66(2):413–457.

Halpern, J. Y., and Pearl, J. 2005. Causes and Explanations: A Structural-Model Approach. Part II: Explanations. *Brit. J. Philos. Sci.* 56(4):889–911.

Halpern, J. Y. 2008. Defaults and Normality in Causal Structures. In *KR*.

Halpern, J. Y. 2014. Appropriate Causal Models and Stability of Causation. In *KR*.

Hansson, H., and Jonsson., B. 1994. A Logic for Reasoning about Time and Reliability. *Form Asp Comp* 6(5):512–535.

John, R. I., and Innocent, P. R. 2005. Modeling Uncertainty in Clinical Diagnosis Using Fuzzy Logic. *IEEE Transactions on Systems, Man, and Cybernetics* 35(6):1340–1350.

Kleinberg, S. 2012. *Causality, Probability, and Time*. Cambridge University Press.

McCarthy, J., and Hayes, P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Machine Intelligence* 4(463-502):288.

Mill, J. S. 1843. *A System of Logic*. Lincoln-Rembrandt.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge University Press.

Reiter, R. 1996. Natural Actions, Concurrency and Continuous Time in the Situation Calculus. In *KR*.