Using machine learning to find causality between errors in log reports

Vithor Bertalan

Polytechnique Montréal
Laboratoire DORSAL

Objectives

- Detect causal relations between log reports.
- If one error is not correctly addressed, if may lead to another errors.
- Therefore, tracking the initial exceptions is fundamental to a good software management.

Objectives

- Detect causal relations between log reports.
- If one error is not correctly addressed, if may lead to another errors.
- Therefore, tracking the initial exceptions is fundamental to a good software management.

Objectives

- Detect causal relations between log reports.
- If one error is not correctly addressed, if may lead to another errors.
- Therefore, tracking the initial exceptions is fundamental to a good software management.

Objectives

- Finding one root problem can avoid sequential ones.
- The development team will have more knowledge to deal with the primordial issue instead of dealing with sequels.
- In the future, we may be able to explain why some errors happened.

Objectives

- Finding one root problem can avoid sequential ones.
- The development team will have more knowledge to deal with the primordial issue instead of dealing with sequels.
- In the future, we may be able to explain why some errors happened.

Objectives

- Finding one root problem can avoid sequential ones.
- The development team will have more knowledge to deal with the primordial issue instead of dealing with sequels.
- In the future, we may be able to explain why some errors happened.

Current step of the research

- Bibliographical review to find the state-of-the-art in log causality.
- Planning of possible approaches to finding causality.
- Searching for the viability of possible datasets.

Current step of the research

- Bibliographical review to find the state-of-the-art in log causality.
- Planning of possible approaches to finding causality.
- Searching for the viability of possible datasets.

Current step of the research

- Bibliographical review to find the state-of-the-art in log causality.
- Planning of possible approaches to finding causality.
- Searching for the viability of possible datasets.

Bibliographical Review

- Most approaches use heuristics to track causality.
  - Aguilera et. al. (2003) use the connections between system calls in their dataset.
  - Mace, Roelke and Fonseca (2018) use Lamport's happened-before relation.
  - Thereska et. al. (2006) use timestamps and breadcrumbs of their reports.

Bibliographical Review

- Recently, some researchers have begun using machine learning to help in this task.
  - Chen, Qi and Hou (2016) use a Bayesian algorithm.
  - Yen, Moh and Moh (2018) use a model combining a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network.
  - Zhou et. al. (2021) use a Hierarchical Attention Network.

Possible approaches

- Use machine learning to track association rules in the dataset, using market basket analysis.

Market basket analysis

- Given a dataset with transactions, we try to find the associations between the items.

Transaction 1	🍎 🍺 ⚪ 🍗
Transaction 2	🍎 🍺 ⚪
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 ⚪ 🍗
Transaction 6	🍼 🍺 ⚪
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

Source of the pictures: https://nextjournal.com/association-rule-mining/filter-association-rules

Market basket analysis

- Then, we find the rules are outweigh our support *(proportion of transactions in which the given item/itemset/association rule holds true)* and confidence *(the proportion of itemsets that contain both X and Y, out of all itemsets that contain X)*.

$$\text{Support } \{🍎\} = \frac{4}{8}$$

$$\text{Confidence } \{🍎 \rightarrow 🍺\} = \frac{\text{Support } \{🍎, 🍺\}}{\text{Support } \{🍎\}}$$

Challenges and future steps

- Continue the review, since significant discoveries in the field have been published in the last few months.
- Finding alternative approaches, in case the association rules method proves ineffective.
- Avoiding overspecialization to a specific dataset.

Challenges and future steps

- Continue the review, since significant discoveries in the field have been published in the last few months.
- Finding alternative approaches, in case the association rules method proves ineffective.
- Avoiding overspecialization to a specific dataset.

Challenges and future steps

- Continue the review, since significant discoveries in the field have been published in the last few months.
- Finding alternative approaches, in case the association rules method proves ineffective.
- Avoiding overspecialization to a specific dataset.

References

- Aguilera, M. K., Mogul, J. C., Wiener, J. L., Reynolds, P., Muthitacharoen, A. (2003). Performance debugging for distributed systems of black boxes. Operating Systems Review (ACM), 37(5), 74–89. https://doi.org/10.1145/1165389.945454
- Chen, P., Qi, Y., Hou, D. (2016). CauseInfer : Automated End-to-End Performance Diagnosis with Hierarchical Causality Graph in Cloud Environment. IEEE Transactions on Services Computing, 12(2), 214–230. https://doi.org/10.1109/tsc.2016.2607739
- Mace, J., Roelke, R., Fonseca, R. (2018). Pivot tracing: Dynamic causal monitoring for distributed systems. ACM Transactions on Computer Systems, 35(4). https://doi.org/10.1145/3208104
- Thereska, E., Salmon, B., Strunk, J., Wachs, M., Abd-El-Malek, M., Lopez, J., Ganger, G. R. (2006). Stardust: Tracking activity in a distributed storage system. Performance Evaluation Review, 34(1), 3–14. https://doi.org/10.1145/1140103.1140280
- Yen, S., Moh, M., Moh, T. S. (2019). CausalConvLSTM: Semi-supervised log anomaly detection through sequence modeling. Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, 1334–1341. https://doi.org/10.1109/ICMLA.2019.00217
- Zhou, C., Li, B., Sun, X., Bo, L. (2021). Why and what happened? Aiding bug comprehension with automated category and causal link identification. Empirical Software Engineering, 26(6). https://doi.org/10.1007/s10664-021-10010-8

Thank You