Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

# A Framework for Anomaly Reasoning: Interpretation through Concept Formation for Knowledge Transfer and Lifelong Learning

**John Winder**
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, Maryland 21250

## Abstract

I am developing a framework for anomaly reasoning for agents that plan and learn in complex, sometimes unfamiliar domains. Anomaly reasoning encompasses recognizing, interpreting, and reacting to unfamiliar objects or familiar objects appearing in unexpected contexts. As a first approach, I propose an interpretation method in which agents form concepts from perceptions to create new representations for use in planning and decision making. An anomaly reasoning framework will make agents more versatile, facilitate learning transfer by pruning irrelevant features, relate new to known phenomena with appropriate similarity metrics, and guide an agent to aspects of the environment most significant to its goals.

## 1 Introduction

In daily life, we encounter novelty of all kinds: unfamiliar words, new sensations, unknown objects, and (most commonly) unusual combinations of recognizable perceptions. Without the capacity to comprehend such spontaneity and its corresponding uncertainty, we would quickly meet with catastrophic failure. Likewise for artificially intelligent agents, the ability to reason about anomalies would improve the speed and quality of learning, minimize failure, and ensure versatility. In this work, I use *anomaly* to describe a novel or unexpected object or occurrence with respect to an agent's purpose, experience, and typical environment. Anomaly reasoning, thus, is an agent's ability to apprehend and suitably incorporate anomalies into decision making, a special case of reasoning about uncertainty.

Consider a robot that has learned the task of place-setting, how to arrange plates, utensils, and chairs around a table. Now suppose that robot enters a room with an anomaly: furniture and tableware of a design, shape, size, and texture it has never perceived before. The robot should still recognize forks as forks and chairs as chairs, and perform its task accordingly, as a human would. This process is a kind of analogical, case-based reasoning over the similarities of an anomaly to prior experiences, interpreting new phenomena by mapping them into an abstract space based on their attributes. An agent could construct and maintain a structured model of its world for use in future decision-making. One such organization of knowledge is a *concept lattice*, which provides a hierarchical relationship among objects and their properties. With this framework, though the new utensils' literal, instantiated values may differ from any known cutlery, their projection into a conceptual space should allow the agent to find an analogical similarity and react appropriately. Anomaly reasoning is not just limited to mapping new objects to known ones; it would help agents recognize and address hazards, atypical environmental properties, and important differences in otherwise familiar objects.

I propose developing a framework for anomaly reasoning. Though broad in scope, anomaly reasoning is decomposable into three distinct sub-problems: *identification* of an anomaly, *interpretation* of it, and *adaptation* to it. Identification is extracting or segmenting discrete objects from an input state and detecting any anomaly. Interpretation projects an identified anomaly into a learned, internal model of the world. Adaptation determines which aspects of the interpreted anomaly are most relevant and what subsequent action is most appropriate to take. Although each of these subproblems is difficult to solve, they can also be addressed separately and integrated into a cohesive pipeline. My primary investigation will center specifically on developing novel methods for interpretation. I hypothesize this goal will necessitate combining a machine learning approach with an organized representation of knowledge. That is, an agent will perceive its world, extract features, and learn while also forming concepts (internal abstractions of experienced phenomena) in a graphical structure that is used to interpret anomalies, for instance, by finding their closest-known analogue. Concept formation, thus, would be an online ground-up construction of a world model in which an agent retains learned objects, attributes, and the relations among them. Together, anomaly reasoning and concept formation help to achieve knowledge transfer for lifelong learning agents operating under uncertainty.

## 2 Approach

To date, I have formalized anomaly reasoning in terms of mathematical functions that take an input state, segment it into objects, detect anomalies among the objects, map them into and update an abstracted concept lattice, highlight relevant attributes, and adapt according to the agent's context by reshaping the state used by their planning or learning algorithm. I have outlined a first approach for anomaly interpretation, which is a natural starting point, since rudimentary forms of identification and adaptation can be provided manually. That is, we can automate feature extraction, tell an agent when an anomaly is present, and have it adapt naively by treating an anomaly as its closest analogue.

The initial approach leverages Formal Concept Analysis (FCA) to provide the knowledge structure. FCA is a mathematical theory for extracting formal concepts (paired sets of objects and attributes) from data, yielding a partial ordering called a concept lattice. Concepts are ordered from the most general supremum (paired set of all objects and empty attribute set) to the infimum (no objects, all attributes). Mapping an object into a lattice gives a sequence of concepts from the supremum to that object's most specific "object concept" (set of objects sharing all its attributes). FCA also permits implication inference, clustering, and developing similarity measures, leading to greater model interpretability. FCA is commonly used for mining static data sets for semantic relations; to my knowledge, this proposed approach would be the first to use FCA interactively for agent-based model-building.

For interpretation, the chosen model is not a lattice but a compressed graph of attribute sets learned and abstracted iteratively from lattices using supervised machine learning. Storing all objects and rebuilding a lattice from them is unfeasible beyond toy domains; abstraction is required but also desirable as it is a form of auto-encoding, in which dimensionality is reduced while distortion is minimized, increasing portability. Error and reduction can be based on heuristics such as concept distance and object support (cardinality of an concept's object set). An anomaly is interpreted as a path in the abstracted graph, in which nodes are attribute sets that can be re-weighted according to an agent's context and are passed on to an adaptation algorithm as a re-representation of the agent's state. Interpretation in this context is closely related to conceptual clustering, case-based reasoning, and representation learning, and in a sense it is a combination of all three. This proposed abstraction process is a novel contribution, and I will investigate the effect of alternate representations including topological orderings, flattened Laplacian matrices, and graph neural networks. Two challenges of FCA are that the data set is recoverable from a lattice and that the number of concepts grows exponentially as more objects are added; the proposed iterative abstraction is a solution to both complexity issues, and should be of interest to the larger FCA community.

I am currently developing a lattice-abstraction prototype to operate on a computer game domain called NetHack, which roughly conforms to the "gridworlds" commonly found in the reinforcement learning literature. An agent's primary goal is to move around the grid and survive until the exit is found. NetHack is promising for developing anomaly reasoning because it includes a wide mix of items, furniture, and enemy types, so an agent will often encounter new objects of these kinds. Much functionality is shared among similar objects such that learning to interact with one informs how to act with others. I would expect an anomaly reasoning agent to learn a graph corresponding to three superclusters (enemies, furniture, items), each with groups for various subtypes, such as food and weapons clusters inside the items supercluster. Implementing the system for NetHack and evaluating whether such clusters do in fact emerge should prove valuable to understanding how well a conceptual model can measure similarity. In the future, I intend to model NetHack in as an object-oriented Markov Decision Process to evaluate the interpretation process in a planning and learning context.

Two other potential domains are the guessing games "Twenty Questions" and "I Spy," where there is some secret object that a player attempts to guess by asking questions. Anomaly reasoning could direct active learning, for which an agent questioner hones its ability to select which attributes are most informative to the identity of the object. For I Spy in particular, anomaly reasoning should make vision-based agents more robust to minor differences, such as familiar objects seen from new angles.

## 3 Related Work

To my knowledge, this thesis is the first attempt to address anomaly reasoning holistically, with a combination of statistical methods and logical formalisms. To that end, relevant work is drawn from a variety of areas: anomaly detection [Chandola *et al.*, 2009], representation learning [Bengio *et al.*, 2013], knowledge graph-based FCA [Ferr, 2015], and conceptual clustering. Much recent work is vision-based, such as learning concept hierarchies [Jia *et al.*, 2013], and robots that learn conceptual models of objects through interactive gameplay [Parde *et al.*, 2015]. Other research influential to this work addresses transfer to novel problems using a knowledge base derived from cross-domain (unrelated) tasks [Bou-Ammar *et al.*, 2015]; this thesis shares the same long term goal of using knowledge transfer to achieve lifelong learning for more general agents.

## References


[Bengio *et al.*, 2013] Y. Bengio, A. Courville, and P. Vincent. Representation learning: a review and new perspectives. *TPAMI*, 35(8):1798–1828, 2013.

[Bou-Ammar *et al.*, 2015] H. Bou-Ammar, E. Eaton, J. Luna, and P. Ruvolo. Autonomous cross-domain knowledge transfer in lifelong policy gradient reinforcement learning. In *IJCAI*, 2015.

[Chandola *et al.*, 2009] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: a survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009.

[Ferr, 2015] S. Ferr. A proposal for extending formal concept analysis to knowledge graphs. In *Formal Concept Analysis*, volume 9113, pages 271–286. 2015.

[Jia *et al.*, 2013] Y. Jia, J. T. Abbott, J. Austerweil, T. Griffiths, and T. Darrell. Visual concept learning: combining machine vision and bayesian generalization on concept hierarchies. In *Advances in Neural Information Processing Systems 26*, pages 1842–1850. 2013.

[Parde *et al.*, 2015] N. P. Parde, M. Papakostas, K. Tsiakas, M. Dagioglou, V. Karkaletsis, and R. D. Nielsen. I Spy: An interactive game-based approach to multimodal robot learning. In *AAAI-15 Workshop on Knowledge, Skill, and Behavior Transfer in Autonomous Robots*, 2015.