# Generalization across Contexts in Unsupervised Computational Creativity

**Debarun Bhattacharjya[1], Dharmashankar Subramanian[1], and Lav R. Varshney[2]**
[1]Research AI, IBM T. J. Watson Research Center
[2]Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
{debarunb,dharmash}@us.ibm.com, varshney@illinois.edu

## Abstract

Autonomous computational creativity systems must not only have the ability to generate artifacts, but also to select the best ones on the basis of some assessment of quality (and novelty). Such quality functions are typically directly encoded using domain knowledge or learned through supervised learning algorithms using labeled training data. Here we introduce the notion of unsupervised computational creativity; we specifically consider the possibility of unsupervised assessment for a given context by generalizing artifact relationships learned across all contexts. A particular approach that uses a knowledge graph for generalizing rules from an inspiration set of artifacts is demonstrated through a detailed example of computational creativity for causal associations in civic life, drawing on an event dataset from political science. Such a system may be used by analysts to help imagine future worlds.

## Introduction

Computational creativity (CC) systems are intended to help generate artifacts that are deemed to be creative by their users or experts in the domain. Although creative value is subjective and often difficult to pin down, there is consensus in the literature that creative artifacts should be novel as well as valuable or useful (Boden 1990; Mayer 1999). Since novelty can be viewed as one of potentially many attributes of creative value (Bhattacharjya 2016), we use the term *quality* to refer to all non-novelty related aspects of creative artifacts (Ritchie 2001; Pease, Winterstein, and Colton 2001).

Clearly, a crucial requirement for any autonomous CC system is the ability to evaluate the creative value of artifacts, particularly their quality. A popular approach to evaluating quality is through the use of extensive domain-specific knowledge. For instance, the IBM Chef Watson system exploits knowledge from hedonic psychophysics around chemical compositions and flavor profiles of individual ingredients (and their combining rules) to evaluate the potential pleasantness of recipes (Varshney et al. 2013). An alternate approach to evaluating quality is to learn it from assessments of artifacts provided by other agents, typically humans. Examples include the PIERRE system for stew recipes that uses human-specified ratings for complete recipes (Morris et al. 2012) and the DARCI system for images that receives feedback from humans (Norton, Heath, and Ventura 2010).

In the aforementioned systems, the mechanism for evaluating quality is explicitly specified; we refer to this as *supervised* computational creativity, whether achieved by supervised learning on complete artifacts or by encoding properties of components and combining rules for combinatorial creativity. In contrast, we posit that quality can be inferred in numerous ways without such explicit knowledge in *unsupervised* computational creativity, using an *inspiration set* (also known as *inspiring set* (Ritchie 2001)) of artifacts and potentially additional knowledge that does not pertain directly to the quality of artifacts. In this paper, we describe a specific data-driven framework that uses a knowledge graph in addition to the inspiration set. We illustrate the approach through a novel application where we first build an inspiration set of cause-effect pairs from a political event dataset and then use these to generate creative cause-effect pairs of events occurring in a country.

The key high-level idea behind our approach is that of *generalization*, i.e. when there is information about artifacts from various contexts, one might be able to learn across all these contexts to estimate a proxy measure for quality. In particular, if there are patterns that seem to be widely prevalent, they could indicate characteristics of high-quality artifacts; the underlying assumption is that widespread prevalence hints at potential usefulness. An artifact could then be contextually creative if it is contextually novel, i.e. original/surprising for a particular context, but also adheres to generally prevalent patterns. While the notion of generalization in computational creativity has been described previously, e.g. Ventura (2016), here we make the connection to quality evaluation – an essential module for any CC system.

Although supervision has played a dominant role in practical CC systems and will likely continue to do so in the future, we believe that the supervised/unsupervised distinction is useful for the field to consider. For one, unsupervised computational creativity forges a path to pursuing abstract conceptual work, thereby enabling ideas and formulations that could be useful across application domains. Further, it explicitly extends the role of machine learning in computational creativity, cf. Guzdial and Riedl (2018). In a recent review, Toivonen and Gross (2015) discuss the role of data mining and machine learning in computational creativity:

as far as evaluation of quality is concerned, they focus on supervised techniques. In contrast, we consider the use of unsupervised learning techniques. We begin by expanding upon the distinction around supervision.

## Supervision in Computational Creativity

There are two fundamental approaches to supervision in computational creativity: 1) to use domain knowledge to map combinations of components of artifacts to measures of quality, and 2) to learn such quality functions from labels such as user ratings, typically through supervised machine learning techniques.

### Quality Functions from Domain Knowledge

Computational creativity applications span diverse application domains such as the visual and culinary arts, music, poetry/narrative generation, and mathematical and scientific discovery. It is not surprising that many successful CC systems rely heavily on knowledge specific to their domain of application, pertaining to the quality of an artifact. This enables the formulation of models that explicitly relate specific combinations of components of artifacts to measures of quality that are appropriate for the application domain. We illustrate this with a couple of examples, one from the culinary arts and one from the sciences.

**Chef Watson.** The Chef Watson system is designed to produce novel and flavorful culinary recipes (Varshney et al. 2013; Pinel, Varshney, and Bhattacharjya 2015). Bayesian surprise, a domain-agnostic information-theoretic notion of novelty, is used together with a combination of two domain-specific measures of flavor quality. The first measure pertains to olfactory pleasantness, drawn from hedonic psychophysics theory and computed from molecular properties of individual flavor compounds present in individual ingredients, together with a combining rule to predict the percept of complete dishes from the individual compound properties. The second is a notion of flavor pairing, drawn from the network science of culinary practice, and was originally validated using positive examples of recipes from large corpora. It is also computed using the flavor compound composition of ingredients. As can be noted, the evaluation of quality in this system requires access to detailed hedonic psychophysics and chemoinformatics data.

**HAMB (Heuristic Autonomous Model Builder).** As a knowledge discovery system that has been deployed in biological sciences applications like protein crystallization (Livingston 2001; Buchanan and Livingston 2004), HAMB is different from other CC systems in that the form of the eventual creative product is different from that of artifacts in the inspiration set. HAMB receives an empirical dataset as input and returns a set of discovery items. These items are varied; the most prevalent kind is a conditional rule that classifies features/attributes based on other features in the dataset (ex: if $f_1$ and $f_2$ then $f_3$ with p-value $p$). The quality of a discovery item in HAMB is its interestingness, quantified using the system builders' expertise around the knowledge discovery process. For example, for a rule or a rule set, it is measured through standard performance metrics for classification such as precision and recall, p-value, etc. HAMB is a prime example of how artifact quality in a CC system is modeled using rich knowledge about the domain, in this case that of rule induction for knowledge discovery.

### Learning from Quality Labels

An alternate approach to supervision in computational creativity is through the availability of what we refer to as *quality labels*. These labels are indications from sources such as previously acquired datasets or real-time human assessments with explicit information about the quality of artifacts. When such labels are available, they can be used to learn one or more quality functions, typically using supervised machine learning methods. Once again, we provide specific examples, from the visual and culinary arts.

**NEvAr (Neuro Evolutionary Art).** In the NEvAr tool, populations of images are generated through an interactive evolutionary process (Machado and Cardoso 2002). Like previous evolutionary art tools, the underlying representation of an image is a tree of mathematical functional operators applied to x-y coordinates of pixels in the image. In NEvAr, the user guides the highly interactive process by selecting individual images and providing a fitness score. Images have a default fitness value of $0$ but the user could choose a small set of preferred images and provide a score greater than $0$, typically $1$ to $3$. This approach is typical of CC systems involving genetic algorithms, where human-assisted supervision is performed in real-time.

**PIERRE (Pseudo-Intelligent Evolutionary Real-time Recipe Engine).** PIERRE is a recipe generation system for crock pot recipes, i.e. soups, stews and chilis, that uses online recipes as an inspiration set (Morris et al. 2012). Like NEvAr, PIERRE uses a genetic algorithm for generation. Crossover is performed by splitting the two parent recipes into two sub-lists each and merging these, and mutation includes changes to ingredient amounts as well as ingredient replacements, additions, and deletions. Supervision in PIERRE occurs through user ratings of recipes which are also available in their repository. A multi-layer perceptron is used to perform a regression that connects an input layer of real-valued amounts of ingredient groups to a real-valued output node of rating (between $0$ and $1$) through a 16-node hidden layer. The system builders also added negative examples by assigning a $0$ rating to randomly generated recipes.

We note that CC systems can be varied and complex in their architecture as well as in the extent and timing of human involvement; this can make it difficult to strictly categorize or contain the mechanism of supervision. CC systems that use case-based reasoning, for instance, could potentially rely on various forms of supervision. An example is poetry generation using COLIBRI (Diaz-Agudo, Gervas, and Gonzalez-Calero 2002) where supervision is achieved by finding the nearest case but also by word substitution using domain knowledge about poetry such as part-of-speech, rhyme, and syllable matching.

Another interesting supervised system is The Painting Fool (Colton 2012) which uses a pipeline of techniques to modify initial domain-specific quality function knowledge. Colton (2008) describes an approach that produces scenes similar to downtown Manhattan where the fitness of the size, shape, color, and location of rectangle placeholders are hand-crafted; an evolutionary model then invents new fitness functions. A practical complication is that a system may work in different modes, perhaps with different types of supervision. NEvAr, for instance, uses quality labels when in interactive evolutionary mode, and author-provided domain knowledge about the aesthetic appeal of an image (based on compression metrics) when in fully automated mode.

A fundamental issue with supervised computational creativity approaches is that it is difficult to transfer quality evaluation modules from one application domain to another. Another issue is that when a system is tied to pre-specified notions of quality, it could miss out on productive regions of the conceptual space of artifacts (Wiggins 2006). Unsupervised techniques could potentially open up the playing field around domain-agnostic quality evaluation in CC systems.

## Unsupervised Computational Creativity

In unsupervised computational creativity, one must attempt to create without the help of an explicit quality function. An approach that is popular is to take an inspiration set of unlabeled positive examples from the domain, learn models to mimic the style and then make modifications of the learned representation. A classic example is the work of David Cope in music creativity, which models the styles of great composers like Bach and Mozart, and then creates new examples of music ranging from single-instrument arrangements to full symphonies (Cope 1996). This approach also allows mixing of two or more different styles.

A modern reincarnation of this approach uses deep neural networks and generative adversarial networks in creative domains, building on their recent successes in machine learning. An example is the work of the Google Magenta project[1] with applications in music and visual art. In certain aspects, this approach to creativity can be limiting. When modifications to the learned representation are minor, resulting artifacts can be perceived to be too close to those in the inspiration set; from an artistic perspective, some have therefore criticized the results as pastiche. When modifications are major, the resulting artifacts may be of low quality, particularly since these systems do not typically have a means to judge their creations. To help avoid such issues, it could be beneficial to use proxies for quality for evaluating artifacts. As Colton and Wiggins (2012) write: “A poet with no critical ability to judge its own work ... is no poet at all”.

Unsupervised computational creativity is clearly a challenging endeavor and necessarily requires making assumptions. This is analogous to machine learning, where unsupervised methods such as clustering implicitly assume that objects that are similar in the feature space are more likely to belong in similar clusters.

[1]https://magenta.tensorflow.org

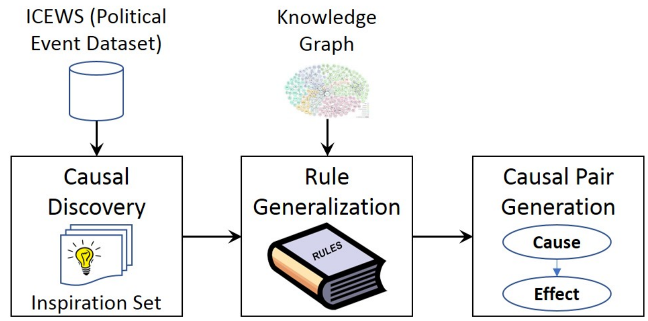



Figure 1: Workflow in the causal association application.

One can further note that due to the absence of any sort of evaluation, a final selection step is often carried out by humans. This is true for Cope’s work but also for many supervised systems, including Harold Cohen’s AARON for visual art (McCorduck 1990).

## Contextual Computational Creativity

We refer to the type of computational creativity formulation underlying the application in the next section as *contextual computational creativity*, where it is assumed that there is access to an inspiration set with artifacts pooled together from various contexts. Formally, the dataset is $\mathcal{I} = \{(z_i, c_i)\}_{i=1}^{M}$ where $z_i$ is the $i$th artifact and $c_i$ is the context of the $i$th artifact, $c_i \in \mathcal{C}$ for some context set $\mathcal{C}$. The contextual inspiration set is the subset that pertains to a particular context $c$, i.e. $\mathcal{I}_c = \{(z_i, c_i) :\ni c_i = c\}$. An artifact in the inspiration set could in general be associated with multiple contexts, and could involve a potentially complex interplay of various constituent components. Examples of inspiration sets of this type include recipe repositories tagged with cuisine information, a database of songs with their genres, etc. In the following section, we present an application of contextual computational creativity that highlights the use of generalization in the unsupervised setting.

## Application: Creative Causal Associations

Analysts in domains such as financial, business, or intelligence analysis are often expected to use their creativity to imagine future worlds. Computational creativity methods could help analysts with divergent thinking, which is an important frame of mind for analyzing long-term and wide-ranging eventualities for scenario analysis (Heuer and Pherson 2010, p. 133). We describe an application in creative causal association that could spark ideas about future events. We explain the steps of our workflow as shown in Figure 1, where a dataset of political events is utilized for generating creative pairs of causally associated events in a country.

### Causal Discovery: Building the Inspiration Set

**Event Dataset.** In relational (also known as dyadic) event datasets, events take the form ‘*who* does *what* to *whom*’, i.e. an event $z$ involves a source actor $a_z$ performing an action/verb $v_z$ on a target actor $a'_z$, denoted $z = (a_z, v_z, a'_z)$.

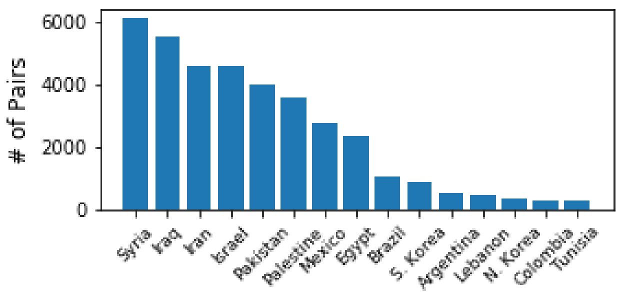



Figure 2: Bar chart showing the number of pairs (artifacts) in the inspiration set for 16 out of the 17 countries in scope. (India with over 45K pairs is omitted from this chart.)

The political science community has been building and curating such datasets for decades; see Schrodt and Yonamine (2013) for a review. While early datasets were obtained through human coding, this has been replaced by automated natural language processing methods that convert news articles in multiple languages into events.

For our application, we use the machine-generated Integrated Crisis Early Warning System (ICEWS) political event dataset (O'Brien 2010), with actors and actions from the Conflict and Mediation Event Observations (CAMEO) ontology (Gerner et al. 2002). Actors in this ontology could either be associated with generic actor roles and organizations (ex: *Police (Brazil)*) or they could be specific people (ex: *Hugo Chavez*). Actions in the CAMEO framework are hierarchically organized into 20 high-level actions and they can be classified by whether they are verbal or material and whether they involve cooperation or conflict.

In our experiments, we restrict attention to events that occurred in India and the 16 countries mentioned in Figure 2 in the time period 1/1/2011 – 12/31/2015. These are primarily countries from Asia and South America, and were chosen to try to find interesting interactions among actors within and across countries. The data was filtered to only include 17 actor roles, including *Citizen*, *Head of Government*, *Protester*, *Insurgent*, etc. Some manual curation was required to transform individuals who are current or former heads of government into their corresponding roles.

**Causal Association.** Causal discovery is a subject of great interest in AI and broadly across the sciences (Pearl 2009). Discovering causal association between a pair of events is typically done through human assessments (Singh et al. 2002) or learned from textual corpora (Radinsky, Davidovich, and Markovitch 2012; Luo et al. 2016). In this work, we have access to a structured *event* dataset of the form $\{(e_k, t_k)\}_{k=1}^{N}$, where $e_k$ is the event type and $t_k$ is the time of occurrence, $t_k \in R^+$. The dataset is strictly temporally ordered with initial time $t_0 = 0$ and end time $t_{N+1} = T$, where $T$ is the total time period. We attempt to discover pairwise causal association by exploiting the fact that an event dataset can be modeled as a temporal point process and therefore represented using a conditional intensity model (Gunawardana and Meek 2016).

We make a simplifying modeling assumption: for a candidate cause-effect pair $(x, y)$, suppose that the intensity of $y$ at any time only depends on whether at least one event of type $x$ has occurred in a preceding fixed window $w$. It can be shown that like the base rate of the effect $\lambda_y$, the conditional intensity parameter $\lambda^w_{y|x}$ can also be computed using summary statistics:

$$\lambda_y = \frac{N(y)}{T}; \lambda^w_{y|x} = \frac{N^w(x \leftarrow y)}{D^w(x)}, \tag{1}$$

where $N(y)$ counts occurrences of event $y$, $N^w(x \leftarrow y)$ counts occurrences where $y$ occurs and at least one event of type $x$ occurs within the preceding feasible time window $w$, and time period $D^w(x) = \sum_{k=1}^{N+1} \int_{t_{k-1}}^{t_k} I^w_x(t)dt$. Here $I^w_x(t)$ is an indicator for whether $x$ has occurred at least once in a feasible window $w$ preceding time $t$.

We propose a causal association score for the pair $(x, y)$ that measures how the conditional intensity of effect $y$ is modified by the presence of potential cause $x$. We refer to this score as the *conditional intensity ratio* with respect to the base rate, $CIR_B(x, y) = \lambda^w_{y|x}/\lambda_y$. We compute these scores for all event pairs for all 17 countries under consideration in one pass each through the country-specific datasets, using window $w = 15$ days and a minimum co-occurence $N^w(x \leftarrow y) = 20$ over the $T = 5$ year time period. We further filter out those pairs in a country whose scores are less than the mean score for that country. This process yields an inspiration set of causal pairs $(x, y)$, counts of which are shown in Figure 2. India has the maximum number of events in ICEWS and ends up with at least one order of magnitude more pairs than any other country in scope.

## Rule Generalization with Knowledge Graphs

There are many approaches to learning general relationships from artifacts in the inspiration set. Here we propose the use of knowledge graphs whenever available and relevant. Knowledge graphs, represented $\mathcal{G}(V, E)$, involve vertices $V$ for entities (such as people, places, objects, etc.) and edges $E$ that represent relationships between entities. Large-scale graphs such as DBPedia, Yago, Freebase, and the Google Knowledge Graph are popular in a host of applications; see Nickel et al. (2016) for a review.

Figure 3 provides a partial knowledge graph for our application, where the vertices include actors from Argentina and Brazil. Consider the following causal pair in the inspiration set for Brazil: *Govt (Brazil) Express Intent To Cooperate Govt (Argentina)* $\rightarrow$ *Citizen (Brazil) Disapprove Govt (Brazil)*. This event pair could potentially be generalized by finding paths in the knowledge graph from every actor in the event pair to the country of Brazil. The bold paths in the figure highlight two paths from *Govt (Argentina)* to *Brazil*, one from the neighbor relationship between Argentina and Brazil, and the other from the fact that they are both in the continent of South America. The resulting rule created from the former path is: *isGovtOf (country) Express Intent To Cooperate isGovtOf (isNeighborOf (country))* $\rightarrow$ *isCitizenOf*

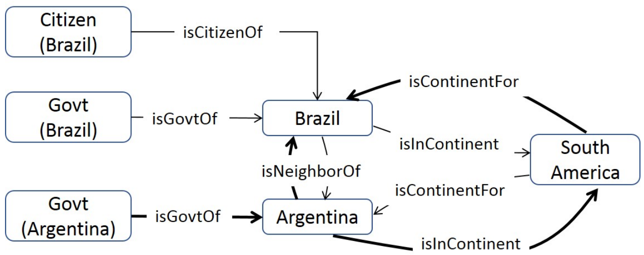



Figure 3: An example partial knowledge graph for selected actors and countries. The two cycle-free paths from *Govt (Argentina)* to *Brazil* are highlighted with bold arcs.

*(country) Disapprove isGovtOf (country)*. Note that the instance from the inspiration set has been generalized and now potentially applies to any country. Similar abstraction paths on graphs pertaining to events have been referred to as predicate projections and have been used for prediction (Radinsky, Davidovich, and Markovitch 2012).

One could proceed in this fashion, compiling rules for all artifacts in the inspiration set $\mathcal{I}$ into a complete list of rules $\mathcal{R}$. We refer to the total number of times a rule $r$ appears in $\mathcal{R}$ as its *support*, denoted $s(r)$.

For our implementation, we constructed an expanded version of the knowledge graph in Figure 3, reproducing similar relations for each of the 17 countries. Aside from the neighboring relation between bordering countries and membership in continents as well as sub-regions (Middle East and South Asia), we also included bi-lateral country relations of alliance (ex: Iran and Palestine) and enmity (ex: India and Pakistan) as they seem particularly suitable for CAMEO coded events of conflict and cooperation.

## Causal Pair Generation

The final stage in our workflow is the generation of creative cause-effect pairs, in which we include the critical aspect of evaluating the quality and novelty of any arbitrary pair.

**Evaluation.** In this unsupervised setting, we estimate quality using the generalization rules. Specifically, if we denote the set of distinct generalized rules satisfied by event pair $(x, y)$ as $\mathcal{R}_{xy} \subseteq \mathcal{R}$, then:

$$q(x, y) \propto \sum_{r \in \mathcal{R}_{xy}} s(r). \tag{2}$$

Thus, our proxy for quality is the total support, which is a measure of how well a causal pair generalizes in aggregate across contexts. Note that according to the proposed metric, specific versions of rules are scored higher than their generalizations.

There are several ways of evaluating the novelty of an artifact in problems of contextual computational creativity. A reasonable approach is to compare the components of the artifact under consideration with those prevalent in the context. In our application, artifacts involve events with actors and actions; we consider an event contextually novel if the frequencies of the source actor, action, and target actor are low in the contextual inspiration set. The novelty of an event pair averages over both events in the pair. Specifically:

$$n_c(x, y) = \frac{g_c(a_x, v_x, a'_x)}{2} + \frac{g_c(a_y, v_y, a'_y)}{2}, \tag{3}$$

$$g_c(a_z, v_z, a'_z) = (1 - f_c^s(a_z))(1 - f_c^v(v_z))(1 - f_c^t(a'_z)), \tag{4}$$

where $f_c^k(.)$ denotes the frequency of the component type $k$ (either source actor $s$, action $v$ or target actor $t$) in events in the inspiration set $\mathcal{I}_c$. The maximum novelty score is 1 and occurs when both events in a causal pair only include actors and actions that are not present in $\mathcal{I}_c$. Other approaches to measuring novelty are possible but not considered here.

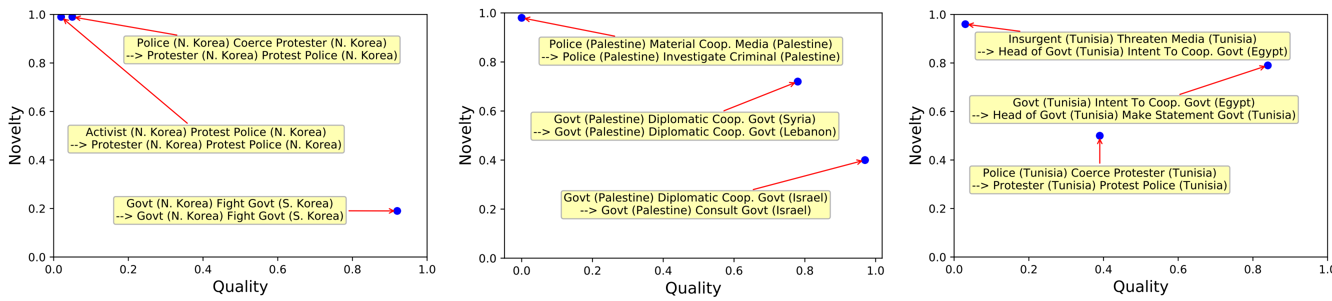
**Generation Methodology.** We generate creative causal pairs for a particular country (context) by first constructing a large set of instances from the complete list of rules $\mathcal{R}$. For every rule, we generate potentially many candidate pairs by traversing backwards on the knowledge graph $\mathcal{G}(V, E)$ from the country under consideration to identify actors along all relation paths in the rule. When a node has many parents that satisfy a particular relation, we randomly choose one of the parents as we walk on the graph. As an example, note that the relation path *isGovtOf (isNeighborOf (country))* from Iraq could lead to *Govt (Syria)* or *Govt (Iran)*. For our experiments, we generate up to 10 unique instances for every rule using these 'random walks', similar to Varshney, Wang, and Varshney (2016).

Once the candidate pairs have been generated, they can be exhaustively evaluated for quality and novelty and then aggregated/ranked in any desired fashion. We normalize quality scores by dividing by the maximum quality pair in a country; novelty is already normalized between 0 and 1.

## Selected Results & Observations

The ranked causal pairs could be used in a variety of ways. For instance, an analyst may wish to review high novelty pairs for assistance in conjuring up future possibilities in a country. Note that by construction, every pair satisfies at least one rule, so there is a minimum quality threshold applied to every pair.

Figure 4 shows selected pairs on a quality-novelty scatter plot from 3 countries in different regions of the world – North Korea, Palestine, and Tunisia. In North Korea (Figure 4 (a)), there is a pair around reinforcement of fighting that is deemed to be high quality as it generalizes well across countries. The two high novelty pairs are perhaps more interesting and involve protests; in one, these are brought on by activism while in the other, protests are caused by police coercion. Recall that we compute novelty using the frequencies of the artifact components in the inspiration set, which in our case are the actions, source actors, and target actors of events. The 10 most frequent components of each type for North Korea are shown in Figure 5. We observe that actions of protest and actors such as protesters and activists are rare in North Korea, which is why they are scored as novel by the system. An analyst may regard large-scale protests in North Korea to be implausible in the near future, yet engaging with

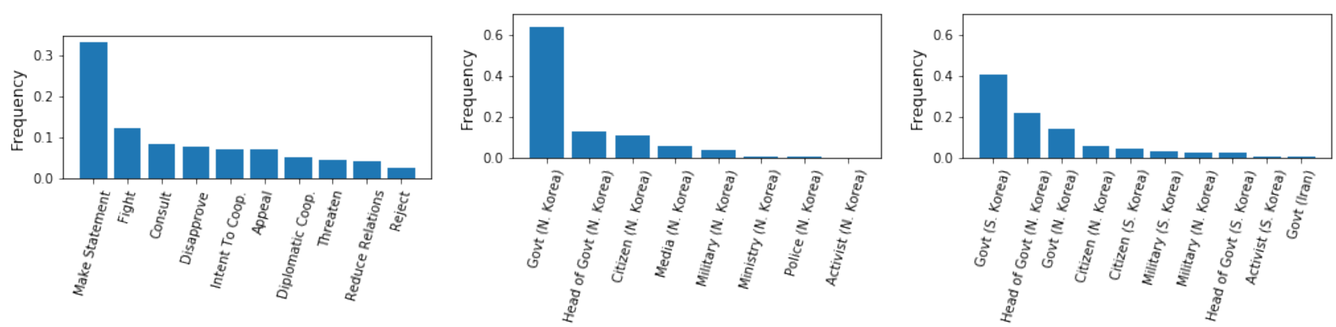



(a) North Korea

(b) Palestine

(c) Tunisia

Figure 4: Selected causal pairs on a quality-novelty scatter plot for 3 countries.




(a) Actions

(b) Source actors

(c) Target actors

Figure 5: Frequencies of the 10 most frequent actions, source actors, and target actors in the inspiration set for North Korea.

the system in this fashion could potentially generate useful ideas and directions for their investigation.

For the sake of comparison, we also plot the pair involving police coercion leading to protests for Tunisia (Figure 4 (c)). A quick investigation reveals that this pair is not as novel in Tunisia since protests are commonplace in the inspiration set. We see instead that pairs where the Tunisian government intends to cooperate with the Egyptian government are deemed novel. Note that Tunisia has the smallest inspiration set (Figure 2).

Due to the focus on country relations in our knowledge graph, we consistently observe high quality pairs involving actions such as diplomatic cooperation, consultation, and the making of public statements across countries; see for instance the high quality pairs in Palestine (Figure 4 (b)). Occasionally, the system is able to identify unusual and quirky pairs, such as the novel pair in Palestine where police cooperation with the media results in a criminal investigation.

We summarize a few other observations from our experimental investigation:

1. The knowledge graph could potentially result in numerous similar instances of the same rule being generated. In our case, group relations such as regional and continental membership result in pairs that appear repetitive as they differ only in the interacting foreign country. This effect could be limited by enforcing further filtering of pairs through additional restrictions and/or by deploying a *variety* module while recommending a set of pairs.

2. Another effect of the choice of knowledge graph, together with the choice of inspiration set, is that generated pairs for a context depend critically on the presence of existing relevant relations. For instance, our omission of USA for this analysis affects Mexico heavily – not having any associations with other countries in our knowledge graph, all of its recommended pairs only involve domestic actors.

3. Due to the aforementioned reasons, the current version of the system does require some human selection, much like other extant systems. The advantage of our proposed approach however is that the system is at least able to self-evaluate artifacts.

We highlight numerous challenges associated with our application. First, building a good inspiration set is difficult because the original data sources are machine-generated and noisy, not to mention the difficulty in discovering causal relations from statistical associations in an event dataset. Furthermore, acquiring and utilizing the appropriate knowledge is essential to the success of learning useful patterns/rules from the inspiration set. The current system is an early foray into work on creative scenarios; we believe that additional progress is required before the system's creations can be usefully evaluated by users.

## Discussion

We discuss how the methods described in the previous section are more general than the application as well as the contextual computational creativity framework that was outlined. We also briefly make connections to a few other relevant concepts in computational creativity.

**Generalizing Generalization (for Evaluating Quality).** In the causal association application, we tried to identify and generate contextually creative artifacts by discovering artifacts that can be deemed novel for a particular context but also satisfy broader relationships learned from artifacts across contexts. Generalizing from the inspiration set could be used to evaluate quality for a broader class of computational creativity endeavors and could therefore be used in other types of applications.

Consider for example the application of creative recipes in culinary art. Varshney, Wang, and Varshney (2016) describe an approach that uses a knowledge graph pertaining to ingredients, which could include information about chemical compounds, seasonality, weather conditions pertaining to ingredient production, etc. One could use techniques similar to those described in the causal association application to generalize from such a knowledge graph along with an inspiration set of recipes, learning rules about which ingredients work well together based on edges (relations) in the graph. Varshney, Wang, and Varshney (2016) do indeed describe an association rule mining approach for learning patterns but they do not make the explicit connection to quality evaluation as we have done here. Association rule mining is one of several potential approaches for generalizing from artifacts that are represented as a set of constituent components, but note that artifacts could be modeled as more complex representations and that relations in such representations could also be generalized in numerous ways.

**Contextual Creativity as P-Creativity.** Boden (1990) distinguishes between *p (psychological)* and *h (historical)* creativity – the former refers to artifacts or ideas creative for a particular individual whereas the latter considers creativity from a historical perspective. In the contextual computational creativity framework outlined here, the intent is to be p-creative in a context by learning from history, through the inspiration set, perhaps along with other knowledge.

**Generalization for Transformational Creativity.** Boden (1990) also makes a distinction with regard to searching for artifacts, referring to producing combinations of familiar ideas and exploring the conceptual space as *combinatorial* and *exploratory* creativity respectively. She regards *transformational* creativity as transforming a conceptual space, such as by adding dimensions or changing constraints.

We highlight that using generalization to evaluate quality could potentially lead to behavior resembling transformational creativity in CC systems, at least in some ways. Injecting new data that is substantially different into the inspiration set could have the effect of modifying the way quality is evaluated and could therefore change constraints during search. Importantly, new knowledge acquired from data sources or other agents could have a more radical effect that alters the way in which quality is assessed.

**A Note on Typicality.** Ritchie (2001) mentions *typicality* of artifacts as another non-novelty related attribute that could be important in a CC system. We have ignored typicality in our application as it is partially built into the generation methodology, like in Morris et al. (2012) – actors that are associated with a particular country can be deemed typical for that context. It may however be useful to incorporate it more explicitly in our application, since one way to remove seemingly redundant cause-effect pairs is to screen out those that seem atypical by only considering a country's frequently associated foreign actors.

## Conclusions

Evaluation is crucial in CC systems since an agent must be able to assess quality. In particular, the assessment function must work for previously unseen artifacts, since novelty is the whole point of creativity. In this paper, we have expounded upon the role that supervision plays in computational creativity by associating it with quality evaluation. Supervision could occur by directly encoding a quality function in a suitably abstract way, but it could also be learned through supervised learning algorithms.

We have proposed generalization as a means to evaluate quality in the unsupervised setting where quality is not specified in any explicit fashion. The benefits of unsupervised generalization in practical CC systems will likely primarily arise when used in conjunction with supervision from other agents. Furthermore, different generalization approaches may be suitable for different types of applications based on artifact and knowledge representations.

The core technical contribution of generalizing with a knowledge graph has been presented in a contextual computational creativity framework, where quality is determined from generalization that borrows strength from artifacts across contexts whereas novelty is context-specific. We imagine that this sort of approach may not be particularly useful when all contexts are similar in the inspiration set, since there would be little capacity to learn something new for any particular context.

We presented a detailed study of an application with cause-effect pairs of political events as artifacts and countries as contexts. Significant work remains towards graduating the proposed techniques in the workflow for the application into a full-fledged CC system. Suitable datasets and better models for causal discovery are essential, aside from improvements in the computational creativity techniques.

## Acknowledgments

We thank our collaborators on the causal association research, Tian Gao and Nicholas Mattei, as well as three anonymous reviewers for their feedback. This work was supported in part by the IBM-Illinois Center for Cognitive Computing Systems Research (C3SR), a research collaboration as part of the IBM AI Horizons Network.

## References


Bhattacharjya, D. 2016. Preference models for creative artifacts and systems. In *Proceedings of the 7th International Conference on Computational Creativity (ICCC)*, 52–59.

Boden, M. 1990. *The Creative Mind: Myths and Mechanisms*. London, UK: Weidenfield and Nicolson Ltd.

Buchanan, B., and Livingston, G. 2004. Toward automated discovery in the biological sciences. *AI Magazine* 25(1):69–84.

Colton, S., and Wiggins, G. A. 2012. Computational creativity: The final frontier. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, 21–26.

Colton, S. 2008. Automatic invention of fitness functions with application to scene generation. In *Proceedings of the EvoMusArt Workshop*.

Colton, S. 2012. The Painting Fool: Stories from building an automated painter. In McCormack, J., and D'Inverno, M., eds., *Computers and Creativity*. Berlin, Germany: Springer-Verlag. 3—38.

Cope, D. 1996. *Experiments in Musical Intelligence*. A-R Editions, Inc.

Diaz-Agudo, B.; Gervas, P.; and Gonzalez-Calero, P. A. 2002. Poetry generation in COLIBRI. In *Advances in Case-Based Reasoning: 6th European Conference on Case Based Reasoning*, 73–87.

Gerner, D. J.; Schrodt, P. A.; Yilmaz, O.; and Abu-Jabr, R. 2002. Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association (ISA) Annual Convention*.

Gunawardana, A., and Meek, C. 2016. Universal models of multivariate temporal point processes. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 556–563.

Guzdial, M., and Riedl, M. O. 2018. Combinets: Learning new classifiers via recombination. arXiv:1802.03605.

Heuer, R. J., and Pherson, R. H. 2010. *Structured Analytic Techniques for Intelligence Analysis*. CQ Press.

Livingston, G. 2001. *A Framework for Autonomous Knowledge Discovery from Databases*. Ph.D. Dissertation, University of Pittsburgh.

Luo, Z.; Sha, Y.; Zhu, K. Q.; Hwang, S.-w.; and Wang, Z. 2016. Commonsense causal reasoning between short texts. In *Principles of Knowledge Representation and Reasoning (KR)*, 421–430.

Machado, P., and Cardoso, A. 2002. All the truth about NEvAr. *Applied Intelligence* 16(2):101–118.

Mayer, R. E. 1999. Fifty years of creativity research. In Sternberg, R. J., ed., *Handbook of Creativity*. Cambridge, UK: Cambridge University Press. 449–460.

McCorduck, P. 1990. *AARON's Code: Meta-Art, Artificial Intelligence and the Work of Harold Cohen*. New York, USA: Freeman.

Morris, R.; Burton, S.; Bodily, P.; and Ventura, D. 2012. Soup over bean of pure joy: Ruminations of an artificial chef. In *Proceedings of the 3rd International Conference on Computational Creativity (ICCC)*, 119–125.

Nickel, M.; Murphy, K.; Tresp, V.; and Gabrilovich, E. 2016. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* 104(1):11–33.

Norton, D.; Heath, D.; and Ventura, D. 2010. Establishing appreciation in a creative system. In *Proceedings of the 1st International Conference on Computational Creativity (ICCC)*, 26–35.

O'Brien, S. P. 2010. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review* 12:87–104.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. New York, NY, USA: Cambridge University Press, 2nd edition.

Pease, A.; Winterstein, D.; and Colton, S. 2001. Evaluating machine creativity. In *Proceedings of Workshop Program of ICCBR-Creative Systems: Approaches to Creativity in AI and Cognitive Science*, 129–137.

Pinel, F.; Varshney, L. R.; and Bhattacharjya, D. 2015. A culinary computational creativity system. In *Computational Creativity Research: Towards Creative Machines*. Atlantis Thinking Machines, Springer. 327—346.

Radinsky, K.; Davidovich, S.; and Markovitch, S. 2012. Learning to predict from textual data. *Journal of Artificial Intelligence Research* 45:641–684.

Ritchie, G. 2001. Assessing creativity. In *Proceedings of the AISB Symposium on Artificial Intelligence and Creativity in Arts and Science*, 3–11.

Schrodt, P. A., and Yonamine, J. E. 2013. A guide to event data: Past, present, and future. *All Azimuth: A Journal of Foreign Policy and Peace* 2(2):5–22.

Singh, P.; Lin, T.; Mueller, E. T.; Lim, G.; Perkins, T.; and Zhu, W. L. 2002. Open mind common sense: Knowledge acquisition from the general public. In *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE)*, 1223–1237.

Toivonen, H., and Gross, O. 2015. Data mining and machine learning in computational creativity. *WIREs Data Mining Knowledge Discovery* 5:265–275.

Varshney, L. R.; Pinel, F.; Varshney, K. R.; Bhattacharjya, D.; Schorgendorfer, A.; and Chee, Y.-M. 2013. A big data approach to computational creativity. arXiv:1311.1213.

Varshney, L. R.; Wang, J.; and Varshney, K. R. 2016. Associative algorithms for computational creativity. *The Journal of Creative Behavior* 20(3):211–223.

Ventura, D. 2016. Mere generation: Essential barometer or dated concept. In *Proceedings of the 7th International Conference on Computational Creativity (ICCC)*, 17–24.

Wiggins, G. A. 2006. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19(7):449–458.