

Searching to Measure the Novelty of Collected Ideas

Rhema Linder
Interface Ecology Lab
Department of Computer
Science and Engineering Texas
A&M University
rhema@ecologylab.net

Andrew M. Webb
Interface Ecology Lab
Department of Computer
Science and Engineering Texas
A&M University
andrew@ecologylab.net

Andruid Kerne
Interface Ecology Lab
Department of Computer
Science and Engineering Texas
A&M University
andruid@ecologylab.net

ABSTRACT

Visual representations of ideas are valuable for creative thinking and expression. Prior research on design and information based ideation has assessed novelty in creative products as the inverse of the frequency that an idea or visual element occurs in the complete space of responses. In controlled experiments, frequency has previously been calculated in reference to the set of ideas collected by all participants (corpus). Experimental conditions restricting the space of possible elements resulted in overlap between participant responses, yielding a range of frequencies. Alas, in field investigations the space of possible elements is unrestricted, resulting in little overlap of ideas, and thus mostly a single frequency ($1/N$) of collected elements.

We introduce a new method that uses web search to measure the novelty of individual ideas. Instead of using the local corpus directly to calculate frequency, we use the number of results from web searches generated by the elements in the corpus. Our implementation uses Google's reverse image lookup to determine the popularity of images. We compare results with those derived via prior methods.

Author Keywords

novelty, information based ideation, evaluation metrics

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Visual representations of ideas are valuable for creative thinking and expression [1, 4]. Prior research on design and information based ideation has assessed novelty in creative products [9, 6, 10].

Information-based ideation (IBI) is the process of having new ideas while working with information [11]. In *information-based ideation tasks*, a person searches, collects, organizes,

and thinks about information to answer open-ended questions, such as planning a vacation, deciding on a thesis topic, or designing an innovation.

In IBI tasks, people author collections of information and media elements from sources including the Web to represent ideas relevant to the task at hand. In this methodology, each collected element is an individual “answer” or idea developed in response to an IBI task, which one can use to measurable components of creativity.

Prior work evaluates *novelty*, one component of creativity, as the rareness of an idea compared to other ideas. Researchers using the creative cognition approach have computed novelty as statistical infrequency in laboratory experiments [9, 10, 11].

We refer to the prior method for calculating elemental novelty as *corpus-based novelty*. In Corpus-based novelty, an element has high novelty if it appears rarely in over the set of all collections in an experiment (the corpus). However, this metric only produces a rich range of values when the number of possible elements an author can collect is constrained by experimental conditions. When users collect media from the internet using self authored queries, the number of possible elements to choose can include any online content. An element collected in such unconstrained conditions is likely unique among participants, but may be commonly or uncommonly found online.

We introduce *search-based novelty*, which does not depend on an a corpus of collected elements from experiments. Instead, we indexed elements from the Internet. While we show empirical evidence to validate this metric in the context of images and Google Image search, we explore potential future work employing search to measure novelty across more media.

PRIOR WORK

Prior work has addressed measuring creativity, ideation process, and the experiential ratings of creativity for evaluating creativity support environments.

Carroll et al. produced a range of standardized questions for evaluating creativity support tools [2]. While user self reported data is important, this work focuses on evaluating the end product participants create to evaluate creativity support environments. Our approach measures creative products, which contain attributes that we can measure, rather than a

Submitted to CHI'13.

Do not cite, do not circulate.

creative processes or experience.

Dow et al. used a variety of metrics to evaluate the effectiveness and creativity of ad prototypes [3]. Instead of coding attributes of the creative products as a measure for novelty, Dow et al. used Mechanical Turk to measure similarity for each ad in the corpus of created ads. To verify, Dow et al. also employed a panel of experts and used the click through rate of ads to measure ad value. Instead of measuring similarity as pairwise, both corpus-based and search-based image metrics described in this paper use statistical infrequency to describe novelty.

Shah et al. review and posit measures and processes for evaluating the effectiveness of methods for generating ideas in the context of engineering design: the intersection of utility and novelty. To assess novelty in the context of a highly controlled experiment, they prescribe aggregating ideas present in experiments into categories. The more often an idea occurs across participants, the less novel it is [9]. Calculating novelty with this approach is easy, but requires many hours of work. Similarly, we use uncommonness to compute novelty, but our analysis is automated.

Webb and Kerne et al. designed a laboratory experiment to validate the hypothesis that automatic organization in an integrated space would improve the creativity of information compositions. *Information composition* is a medium that affords collecting and organizing ideas as text and image bookmarks, helping people perform IBI tasks. Participants collected images from a preselected set of search queries. Under these conditions, nearly all of the images collected were collected by multiple participants, creating significant overlap [11]. We report a field study where each participant collected images with little overlap.

Webb and Kerne use Equation 1 to calculate the image novelty of an individual element image e in a set of all image collections C is 1 divided by the number of sets in C where e is present [11]:

$$enov_i(e, C) = \frac{1}{|c \in C : e \in c|} \quad (1)$$

As elements in a corpus become more common, the value of $enov_i(e, C)$ becomes smaller, indicating that the element is less novel. Also note that the novelty of an element e is dependent on the corpus of participant respondent collections C . Calculating the inverse popularity of elements among a corpus closely follows principles that make IDF useful in information retrieval.

Inverse Document Frequency (Equation 2) highlights rare terms in documents by giving high weights to terms that occur in fewer documents over a corpus [8]. The IDF of a term t in a corpus of documents D depends on the number of times the term appears in a given document d :

$$idf(t, D) = \log \frac{|D|}{|d \in D : t \in d|} \quad (2)$$

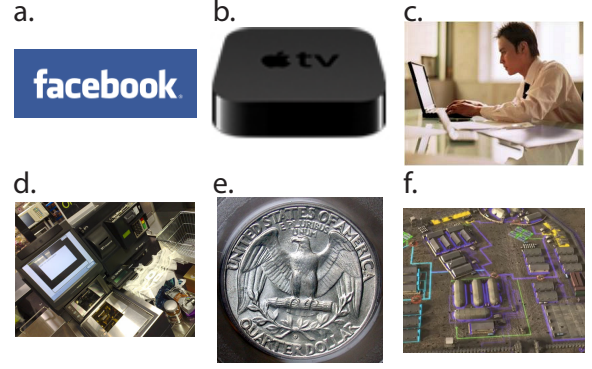


Figure 1. Example images from least to most novel. Images a-f have $inov$ scores of 0, 2, 4, 6, 8, and 1 respectively. The Facebook logo is the least novel. Next, an image of Apple TV2, then a stock photo, a photo of a self checkout system, a us quarter back, and a diagram of power grid.

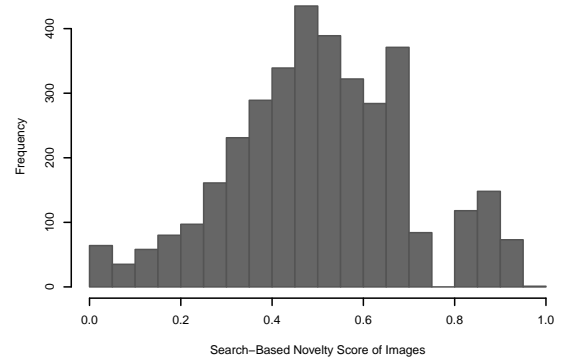


Figure 2. Histogram of the the search-based image novelty ($nov_i(e)$) for 3,579 image queries.

SEARCH-BASED NOVELTY

Search-based novelty measures the uniqueness of an element with the number of pages that element appears on the Web. The process for implementing a search-based novelty metric includes four steps:

1. Collect a number of examples.
2. Generate search queries for each example.
3. Perform searches with generated queries.
4. Create a function to normalize the number of results.

Search-Based Image Novelty

We present our implementation, *search-based image novelty*, which uses Google Image searches to measure the novelty of images found on the Web.

Google Images indexes almost all images on the Web, allowing one to search for images using text or image queries. Google Images combines very similar images, making the search invariant to resolution and small visual perturbations.

We first collect example images C (1). We use the set of 3,579 images from the information compositions that students in

DPCE collected in a fields studies. For each image, we generated a search query (2) using each image’s url as the query. For each image in C , we performed a Google Image search to get the number times the images appears on the Web e_n (3). The number of returned search results determines popularity (e_n).

To create a function that normalizes e_n (4), we take the sum of the logarithms of the popularity of images from all compositions to derive average popularity (\bar{e}). We then double average popularity (\bar{e}) to estimate a maximum, and calculate the normalized image novelty ($inov$) of an element (e). The max we use effectively sets a maximum popularity, which avoids losing resolution because of images in the long tail.

$$\bar{e} = \frac{\sum_{e \in C} \log(e_n)}{|e \in C|} \quad (3)$$

$$inov(e) = 1 - \max(1, \frac{\log e_n}{2 \cdot \bar{e}}) \quad (4)$$

Thus, $inov(e)$ is a function for novelty that does not depend on a corpus of collected responses. \bar{e} does not need to be calculated for every experiment. We found $\bar{e} = 5.02$, to get the resulting function $inov(e) = 1 - \max(1, \log(e_n)/10.04)$. $inov$ will always be between 0 (not novel) to 1 (very novel). We show the histogram for $inov$ scores over C in Figure 2. In Figure 1, we show example images where the Facebook logo is the least novel.

FIELD STUDY

Laboratory experiments help researchers focus in on factors that contribute to creativity by providing a controlled environment. In contrast, investigations in the field provide ecological validity. We conducted a field study in The Design Process: Creativity and Entrepreneurship (DPCE), an interdisciplinary undergraduate course.

DPCE students used our creativity support tool, InfoComposer [5], to author information compositions on soft innovations – new ideas formed from combining and extending existing ideas. An information composition is a holistic visual representation of a collection of image and text bookmarks [6]. Students searched the Internet, collecting relevant image and text bookmarks to represent ideas about their soft innovations. Students’ search queries and potential information sources were *unrestricted*.

To calculate the novelty of a composition, which contains a set of images, we first calculate novelty for each of its images and then calculate a mean. We compare the prior corpus-based image novelty (Equation 1, Figure 3) and the new search-based image novelty (Equation 4, Figure 3).

Corpus-Based Image Novelty Results

Corpus-based image novelty yielded a small range of values, leaning heavily toward high novelty. Figure 3 illustrates how poorly the metric is suited for the unrestricted field study conditions, showing a value of 1 for most compositions. Such extreme values have do not contain enough granularity for valid comparison.

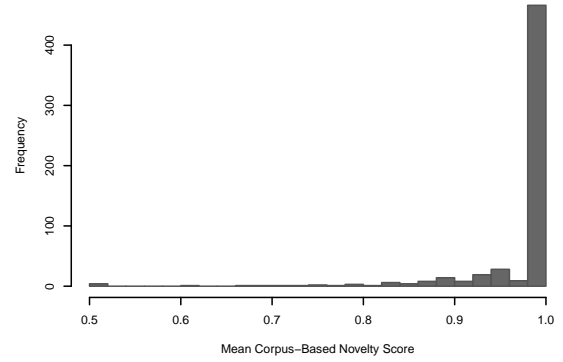


Figure 3. Histogram of corpus-based image novelty scores of 682 student authored compositions from DPCE. Even with the large number of student made compositions, the lack of overlap creates a small range of image novelty scores.

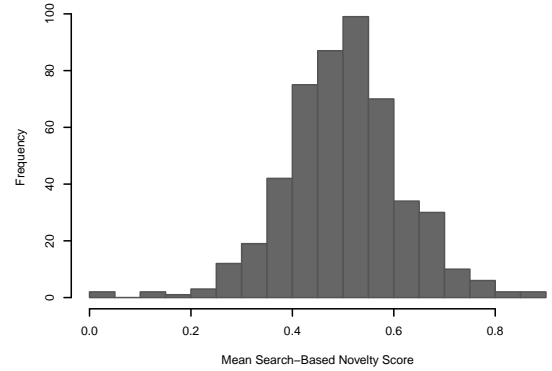


Figure 4. Histogram of search-based image novelty scores of 682 student authored composition from DPCE. Using the same images used in Figure 3, search-based novelty produces novelty scores with more granularity.

Search-Based Image Novelty Results

Search based image novelty scores provided a higher range of values. These values are what one would expect for novelty, showing a range of values. Avoiding ties in values makes this metric better suited for statistical test that require interval values, which are useful for comparing creativity support environments in filed studies and laboratory experiments.

DISCUSSION

In this section, we show how search-based methods for measuring the novelty of elements could be applied to other kinds of media. We discuss the benefits of using automated metrics to analyze products made with creativity support environments.

More Media

We have shown search-based novelty to be more effective for describing image novelty than corpus-based novelty when images can be collected from anywhere on the Web. While we have only implemented and tested search-based image

novelty, we suspect that similar implementations would work with other kinds of media. Once one can generate queries and perform searches for an element of an arbitrary media type, then implementation of a search-based novelty score is straight forward. We envision this metric being used to assess elemental novelty for collections of various media including, audio, text, or video. For text, one could perform searches using samples of extracted text. For audio, one could search using queries from tags.

Aggregating Novelty

Justifying the methods for aggregating elemental novelty into a single score is difficult. We have used the mean of each element to represent the novelty of a collection, but we do not have a clear understanding of which aggregation would be best. As researchers, we hope to find significant differences between conditions in experiments. Transformations designed to accentuate difference could be used on these metrics, but should not be used without theoretical justification.

Elemental novelty does not consider the context of a collection, ignoring relationships between elements. If a collection includes a very popular image among moderately novel images, the overall novelty of the collection may be low. If an author creates novel combinations of ideas with common elements, they will still score low on elemental novelty.

We continue to develop human rated metrics that measure holistic components of creativity in information compositions [7]. While human rated metrics take longer to calculate than automated processes, they can be used in small experiments. Using crowdsourcing services can help mitigate the time any one individual has to spend to rate aspects of collections.

Implications for Design

We have shown that metrics that work well in a laboratory do not always scale well in the context of field studies. Big data is a resource that can be used as a baseline for the originality of content. If researchers can uncover properties that leverage publicly available web services to analyze media, they can use those processes to measure the effectiveness of tool that help people create media rich products. Combining empirical results with experiential data ensures that both the experience and the product of creativity are analyzed.

Statistical methods on computational processes for evaluating creative products help researchers analyze more data because they do not require human analysis. Computational methods can be easier to verify and repeat across research groups. Evaluation at web scale is of interest, because it involves larger sets of users, which are essential to investigating the validity of at least some creativity support environments, such as large university courses, and massively open online courses. Such large scale deployments of creativity support environments necessitate techniques that enable equally large scale automatic assessment of ideation metrics.

Conclusion

We used observations from experimental and field study data to show that search-based image novelty measures are more

effective when the corpus of collected images yields few commonly collected images. By leveraging Google Images, we computed a baseline for image novelty to construct a useful normalized novelty metric. We presented directions for future work that extend this metric to other media types that leverage search to measure novelty.

REFERENCES

1. Arnheim, R. *Visual Thinking*. UC Press, 1972.
2. Carroll, E. A., Latulipe, C., Fung, R., and Terry, M. Creativity factor evaluation: towards a standardized survey metric for creativity support. In *Proceedings of the seventh ACM conference on Creativity and cognition*, C&C '09, ACM (New York, NY, USA, 2009), 127–136.
3. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., and Klemmer, S. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2807–2816.
4. Glenberg, A. M., and Langston, W. E. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language* 31, 2 (1992), 129 – 151.
5. IEL. Infocomposer, <http://ecologylab.net/infocomposer/>, 2013.
6. Kerne, A., Koh, E., Smith, S., Choi, H., Graeber, R., and Webb, A. Promoting emergence in information discovery by representing collections with composition. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, C&C '07, ACM (New York, NY, USA, 2007), 117–126.
7. Kerne, A., Webb, A. M., Linder, R., Smith, S. M., Moeller, J., Damaraju, S., Lupfer, N., and Qu, Y. Components of information-based ideation productivity. *submitted to ToCHI* (2013).
8. Salton, G., and Buckley, C. Term-weighting approaches in automatic text retrieval. *Information processing & management* 24, 5 (1988), 513–523.
9. Shah, J. J., Smith, S. M., and Vargas-Hernandez, N. Metrics for measuring ideation effectiveness. *Design Studies* 24, 2 (2002), 111 – 134.
10. Shah, J. J., Smith, S. M., Vargas-Hernandez, N., Gerkins, D., and Wulan, M. Empirical studies of design ideation: Alignment of design experiments with lab experiments. In *Proc ASME Conf on Design Theory and Methodology* (2003), 1–10.
11. Webb, A. M., and Kerne, A. Integrating implicit structure visualization with authoring promotes ideation. In *Proc. JCDL* (2011), 203–212.