Characteristics of Pro-c Analogies and Blends between Research Publications

[1]Yalemisew Abgaz, [1]Diarmuid P. O'Donoghue, [2]Donny Hurley,
[3]Ehtzaz Chaudhry, [3]Jian Jun Zhang

[1] Department of Computer Science, Maynooth University, Maynooth, Co. Kildare, Ireland.
Diarmuid.ODonoghue@nuim.ie
[2] Department of Computing, Institute of Technology Sligo, Ash Lane, Sligo, Ireland.
[3] National Centre for Computer Animation, Bournemouth University, Bournemouth, UK.

Abstract

Dr Inventor is a tool that aims to enhance the professional (Pro-c) creativity of researchers by suggesting novel hypotheses, arising from analogies between publications. Dr Inventor processes original research documents using a combination of lexical analysis and cognitive computation to identify novel comparisons that suggest new research hypotheses, with the objective of supporting a novel research publication. Research on analogical reasoning strongly suggests that the value of analogy-based comparisons depends primarily on the strength of the mapping (or counterpart projection) between the two analogs. An evaluation study of a number of computer generated comparisons attracted creativity ratings from a group of practicing researchers. This paper explores a variety of theoretically motivated metrics operating on different conceptual spaces, identifying some weak associations with users' creativity ratings. Surprisingly, our results show that metrics focused on the mapping appear to have less relevance to creativity than metrics assessing the inferences (blended space). This paper includes a brief description of a research project currently exploring the best research hypothesis generated during this evaluation. Finally, we explore PCA as a means of specifying a combined multiple metric to detect comparisons to enhance researchers' creativity.

Introduction

Analogical thinking was a frequent mode of thought for eminent scientists like Faraday, Maxwell and Kepler. This paper concerns an analogy-based model to enhance the creativity of practicing scientists by employing a computational model of analogy to uncover novel and potentially useful comparisons between research papers. In order to support its users, Dr Inventor (O'Donoghue, Abgaz, Hurley, & Ronzano, 2015) generates analogy-based comparisons that are similar to those developed by scientists - were they to explore the same comparisons "manually". This objective is achieved through a realistic computational simulation (Gentner & Smith, 2012), building upon several decades of focused work on analogical comparisons and conceptual blending (Fauconnier & Turner, 1998). Computational modelling of analogy has relied primarily on human constructed data (O'Donoghue & Keane, 2012), but Dr Inventor outlined in this paper uses raw data sourced directly from existing research publications. This paper focuses on identifying the most creative comparisons for a given target, so the user only explores the most *Pro-c* creative (Kaufman & Beghetto, 2009) hypotheses.

Analogical comparisons can make a problematic concept seem more familiar, but can also make familiar ideas seem novel and fresh by comparison to some unexpected source. Novel and potentially creative comparisons can highlight previously overlooked facets of the original concept, bringing to light such information. Searching for novel analogies might be one specific mode of the divergent thinking associated with creativity. Potential applications of the Dr Inventor system presented in this paper range from creativity assistant to plagiarism detection (Hurley, Abgaz, Ali, & O'Donoghue, 2016). Thus, the main objective of this paper is to identify qualities and metrics that support the accurate identification of analogous pairs of publications that have the greatest impact on users' creativity.

This paper begins with a review of related work and some background on analogy and conceptual blending. We outline the Dr Inventor model before presenting users' evaluations for a collection of inter-publication comparisons. Several theoretically motivated metrics are statistically examined as predictors of creativity ratings. We also briefly outline a research project that arose from one of Dr Inventor's research hypotheses.

Background and Related Work

The IBM Watson (Pinel & Varshney, 2014) cognitive computing system incorporates a deep parsing of natural language documents, enabling some recent forays into culinary creativity as IBM Chef Watson. While Dr Inventor and IBM Watson both use deep parsing, only Dr Inventor focuses on analogical comparisons (and conceptual blends). Goel, et al. (2015) discuss how students used Watson first as an aid to co-creativity and subsequently as a means of actively enhancing co-creativity.

KnIT (Spangler, et al., 2014) aims to predict scientific discoveries by analyzing past literature. It extracts and collects information from multiple publications, looking for literal similarities that focus on central topics. KnIT has proposed a novel and testable hypothesis related to a tumor suppressing protein called *p53*. While Dr Inventor and KnIT focus on scientific literature, only Dr Inventor explores non-literal similarities between publications.

Literature Based Discovery (LBD) (Bruza & Weeber, 2008) is also arguably a creative undertaking, whose ABC model seeks knowledge (B) connecting distinct bodies of literature (A and C). CrossBee (Juršič, Bojan, Tanja, & Lavrač, 2012) adopts an LBD-like approach by identifying cross-context terms (not documents) that form connections between publications in distinct areas of research.

Dr Inventor differs from IBM Watson and KnIT by focusing on analogical similarities between ideas. Gentner (1983) distinguishes four categories of similarity (Table 1), highlighting differences between surface features and the deep structure of information. Dr Inventor searches for non-obvious structure-based analogies between publications with few obvious surface similarities (few similar objects).

Table 1, *Dr Inventor focuses on Analogical Similarity while deliberately avoiding literal similarity*

	Similar attributes and objects	Similar relational structure	Example
Literal similarity	Many	Many	*Proxima Centauri is like the Sun*
Surface similarity	Many	Few	*A candle is like the sun*
Analogy	Few	Many	*The atom is like the solar system*
Dissimilar or Anomaly	Few	Few	*The atom is like a chicken*

COINVENT (Schorlemmer, et al., 2014) is another concept invention system which attempts to build a formal model of conceptual blending by drawing various interdisciplinary research results. COINVENT is aimed at gaining a deep understanding of conceptual blending and developing a formal method for building a generic creative computational system. COINVENT uses mathematics and music as a working domain, while Dr Inventor focuses on analogical (not literal) comparisons between graphics publications.

Reasons for focusing on analogical similarity (over literal similarity) include: a long-standing view that analogy is an important mode of scientific creativity (Koestler, 1964), existing tools already support literal (but not analogical) similarity, and computational advances in language processing and analogy modelling enable efficient identification of analogies between text-based publications.

Dr Inventor and SIGGRAPH

Dr Inventor is a Creativity Enhancement Tool (CET) operating as a partial simulation of scientists' creative thinking about research literature. Testing uses a corpus of 1146 papers from the SIGGRAPH[1] conference series[2] (2002 to 2015). It adopts a task-divided (Kantosalo & Toivonen, 2016) approach to co-creativity, where Dr Inventor searches for creative sources, but evaluating the *usefulness* of its hypothesis (discussed later) is the users' responsibility.

Reasons for focusing on a specialized domain like computer graphics are, firstly, it ensures a somewhat consistent degree of novelty between publications, allowing any resulting comparisons to be consistently evaluated. Secondly, the resulting inferences should generally be more semantically plausible than might arise from two semantically unrelated papers (say economics and computer graphics). Finally, we may use expert evaluations under the consensual assessment technique (Baer & McKool, 2009), while minimizing any impact from relative expertise in other disciplines.

Previously (Abgaz, et al., 2016-b; Abgaz, O'Donoghue, Smorodinnikov, & Hurley, 2016) explored several metrics measuring the outputs generated by Dr Inventor. This paper builds on that work by exploring new metrics not previously discussed, including several new metrics for mappings and inference that were not previously addressed.

Professional Level *Pro-c* Creativity

Dr Inventor models professional creativity (Pro-c) (Kaufman & Beghetto, 2009) representing "*effortful progression beyond little-c that represents professional-level expertise in any creative area*." Dr Inventor is a model of analogy-based professional creativity, which might never achieve the eminence associated with Big-C creativity.

Boden's (2004) refers to the "three main types of creativity – *combinational*, exploratory, and transformational"[3]. This paper explores one form of combinational creativity, arising from analogies within the SIGGRAPH corpus. We examine combinational creativity and the space of possible mappings. The central graph-matching problem is NP-hard (Veale & Keane, 1997), with the number of potential target-to-source combinations increasing exponentially with the size of each paper. Multiple source papers each represent a new graph matching problem.

Next, we look at combinational creativity and the inferences. Dr Inventor's lexical phase identified 26,072 distinct concepts represented as graph nodes (see formal presentation in (Abgaz, et al., 2016-b)) and 7,315 distinct relations (discounting repeated instances), allowing the creation of $4.9*10^{12}$ distinct combinations of concepts and relations. Furthermore, exploring clusters of just 5 inferences could form $3*10^{63}$ possible inferences, while the presence of co-references increases the space of possible inferences yet further. By comparison there are around $4*10^{79}$ atoms in the observable universe.

We believe Dr Inventor also addresses the issue of intentionality (Ventura, 2016) as it only selects those compari-

[1] ACM Special Interest Group on Computer Graphics and Interactive Techniques

[2] SIGGRAPH is the 5th ranked (of several thousand) conference in computer science by Microsoft Academic Search (accessed February 15, 2017).

[3] My emphasis on "combinational", from Preface to the 2nd Edn.

sons likely to be adopted by an expert user. This paper explores multiple facets of these comparisons to identify those of greatest creative impact. Surprise is also associated with creative comparisons and many of its analogies were found to be surprising, comparing papers from different subtopics that were separated by many years. Dr Inventor's research hypothesis (page 6) identified "hole" as analogous to "area", though these terms seem more like opposites.

While Dr Inventor explores combinational creativity, how can it find those analogies that will have the greatest creative impact on professional users? This central question motivates this paper, leading to our initial hypothesis that it is the mapping (or counterpart projection) that is the hallmark of creative comparisons.

The Dr Inventor Computational System

The Dr Inventor system aims to simulate one mode of creative scientific thinking to identify comparisons that might enhance a researcher's creativity. Dr Inventor achieves this through its deep processing of natural language retrieved directly from research publications, from which it derives an attributed relational graph called a Research Object Skeletons (ROS) in the form *(relation (subject, object))*. Crucially, multiply referenced items are uniquely represented in a ROS (see Figure 1).

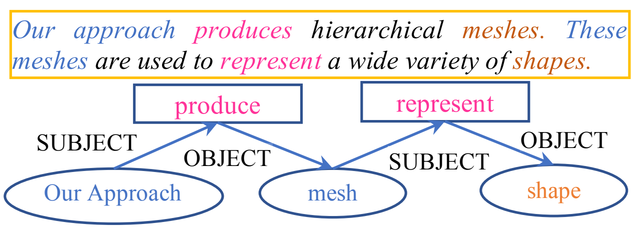

Figure 1. *Subject-verb-object triple generated by the graph builder.*

Unlike human generated data that neatly segregates mapping data from inference data, our ROS involve a single unified knowledge structure identifying the best mapping between any pair of papers (the Largest Common Subgraph NP-hard problem), typically identify different subsets of these ROS, depending on the particular papers being compared. Any un-mapped information from the source ROS thus becomes available as candidate inferences for possible transfer to the target. A semantic grounding constraint ensures that all accepted inferences overlap, at least in part, with the corresponding mapping. This ensures the inferences relate directly to the identified analogical similarity that effectively forms the justification for those inferences. We believe that identifying the pre-existing similarities and detecting potentially transferrable knowledge should be seen as central parts of the creative challenge - and differentiating between them should not be assumed as part of the inputs to the creative process.

Dr Inventor Architecture: Dr Inventor combines a number of key technologies beginning by extracting the text from a publication in PDF format using the tools PDFX and Grobid, addressing multi-column layouts, headers, footers, tables etc. Extracted text is passed to the GATE dependency parser (Ronzano & Saggion, 2015), tailored to deal with inline citations and identifying co-referent terms (like "it" and the item it references). Parsing results undergo semantic extraction to identify key information, generating the ROS graphs that represent each publication through the remainder of the cognitive model. Although SIGGRAPH publications frequently include mathematical expressions these formulae are not currently parsed. However, mathematical variables frequently contribute to mappings due to their use throughout documents.

The analogical mapping between two publications are generated from the ROS graphs using a tailored version of the VF2 (Cordella, Foggia, Sansone, & Vento, 2004) subgraph matching algorithm. VF2 ensures an appropriate balance of two often competing influences, the mapping process of matching semantically similar concepts while simultaneously respecting the different topologies of the input ROS graphs. It restricts the space of possible mappings, allowing *concepts* (noun) to only map with other concepts while *relation* (verb) nodes also map together.

Building on the mapping, Dr Inventor generates the corresponding inferences, representing the expected information created in response to the comparison. Dr Inventor then blends the new information into the pre-existing target and presents it to the user by placing the inferences in the context of the target paper.

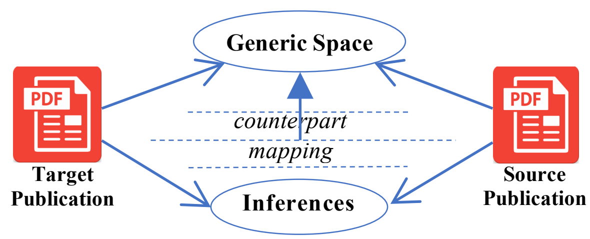

Figure 2: *Conceptual spaces used by Dr Inventor, focused on structure driven counterpart mappings*

Consistent Terminology

An analogy is a structure-based comparison between two collections of information, centered on a *1-to-1 mapping* between these systems (Gentner, 1983). Later this paper investigates characteristics often associated with the more general cognitive theory of Conceptual Blending (Fauconnier & Turner, 1998). However, the generality of blending can blur the distinction between analogy and literal similarity (Table 1). For simplicity, we define some terminology to clarify that all comparisons discussed in this paper involve structural similarity (and not literal or surface similarity). Where relevant, we use the term *counterpart mapping* to indicate the structure-based 1-to-1 mapping between two ROS conforming to structure mapping theory (Gentner, 1983). This we see as a refinement on blending's general concept of the counterpart projection.

A mapped pair $P=(S, T)$ is a tuple of source *(S)* and target *(T)* items (concept or relational nodes) from the two texts.

The mapping consists of paired items, each identifying the non-literal similarity between them, often being taxonomically related and indicating the Generic Space (Figure 2).

Analogy can be a highly profligate inferential process. To guard against generating many unwarranted inferences, Dr Inventor generates only "grounded inferences" that build directly on information contained in the underlying mapping.

Creative Qualities and Dr Inventor

A user evaluation was undertaken to gather ratings for three qualities of creativity, these being selected from the SPECS list (Jordanous, 2012) as being of greatest relevance to scientific creativity (Abgaz, et al., 2016-b).

Participants: 15 experts in computer graphics were recruited between senior Professors, with many SIGGRAPH publications, to postdoctoral researchers and PhD students having the least experience.

Materials: 10 target publications were chosen using stratified random sampling from across the years of the SIGGRAPH corpus. Dr Inventor explored each possible analogy identifying the best source analog for each target. These analogies were selected based on structural and semantic metrics focused on the counterpart mapping, using the *AnaSim* metric described below.

Evaluators were shown a training video describing analogical comparisons and showing use of the Dr Inventor system. The 10 selected analogies were given to the users who spent approximately 50 minutes exploring each publication before giving their rating evaluations. All evaluations were completed using the Dr Inventor system online[4] over the course of three days.

Procedure: Users were provided with electronic copies of the source and target paper on the Dr Inventor system. They were asked to read each paper then were asked to explore the identified analogy, a process that was directly supported by presentation of paired terms from the source and target papers. Evaluators were provided with three alternative visualizations of the mapping to support their understanding, allowing users to navigate between the mapped terms and their locations within the documents.

After exploring each comparison user evaluations were collected, also online via a form embedded into the Dr Inventor system. At least 2 user ratings were gathered for each of the following SPECS inspired questions.

1) This is a Novel or Unexpected comparison

2) This is Potential Useful and Recognizes Gaps in the research

3) This comparison Challenges the norms in this discipline

Inter-Rater Agreement: We investigated the agreement between raters for each of these qualities. Our ordinal data coupled with multiple raters required use of Krippendorff (Krippendorff, 2011) inter-rater agreement, returning values between 0.0 and 1.0 with 1.0 indicating maximum agreement. Krippendorf's alpha for each quality was found to be: *Novelty*= 0.382, *Usefulness*=0.26 and *Challenge the norms*=0.39. While this level of agreement may appear low, we argue that these creativity ratings are still valid firstly because of the relatively large number of rating categories (5), reducing the alpha score. Additionally, creativity is often seen as highly personal and dependent upon users' expertise and experience. Post-evaluation discussions highlighted why experts gave very different ratings for a few comparisons, focusing on expertise on specific topics. We note that *Usefulness* showed the lowest agreement with differences in expertise causing disagreement.

Interestingly, more senior users (Professors and Senior Lecturers) found greater value in Dr Inventor than less experienced researchers. Due to the lack of agreement between raters and the fact that each comparison was the best of 1146 (computationally selected) possible comparisons for that target, the normal distribution was seen as inapplicable. Thus, the following evaluations rely on non-parametric statistical techniques.

Counterpart Mapping Metrics

To quantify the degree of similarity existing between any analogous pair, we employed a number of computational metrics focusing on different aspects of a comparison. First, we specify metrics related to the topological similarity, then metrics for semantic similarity and finally a number of combined metrics incorporating both influences. Metrics presented in this section focus on the counterpart mapping, while the semantic similarity relates to the generic space. Both structural and semantic factors form central parts of the VF2 (Cordella, Foggia, Sansone, & Vento, 2004) based (sub-)graph isomorphism mapping algorithm. Additionally, the semantic scores outlined below directly influence the best mapping that is identified between analogs. As noted by (Van Mieghem, 2013) and others there is no single metric that can usefully compare different graph topologies, leading to our multi-parameter investigation.

Of course, any successful metric might be easily adopted as a basis for data mining (Toivonen & Gross, 2015) for the expeditious identification of creative analogies. However, the focus in this paper is on identifying qualities and associated metrics that perform best as indicators of creative comparisons. Martins, Pollak, Urbancic, & Cardoso (2016) discuss general optimality principles for conceptual blending, but do not address creative comparisons.

We now examine each of the metrics to assess its impact on the level of creativity attributed to each analogy, beginning with the topology-based metrics before moving on to the semantically focused ones.

1. Size of the Mapping (*MapSize*): We examined the result for an expected relationship between the size of the mapping and the ratings awarded by users to each comparison. A Spearman rank order correlation $r_s = 0.212$ (p=0.279) indicated that mapping size was not an influence in users' perception of creativity. This was somewhat surprising as larger mappings were expected to be more convincing and thereby promote creative thinking.

[4] Available from DrInventor.eu

2. Ratio of Mapped Information (*MapRatio*): This counterpart metric quantifies how much of the target graph participates in the mapping. This measure will result in a higher similarity score for targets that have been thoroughly accounted for by the mapping.

$$\text{MapRatio} = \text{MapSize}/\text{TargetSize}$$

MapRatio produces values between 0 and 1 where values near 1 indicate a greater portion of the target problem participates in the mapping.

A Spearman rank order correlation between MapRatio and the average creativity score was calculated but was found to be not significant r_s=-0.006 (*p*=0.492). So MapRatio does not appear to be an important factor in identifying creative comparisons. Again, a somewhat surprising result suggesting that merely re-interpreting a target is not sufficient to cause a creative impact. Of course, MapRatio does not incorporate any measure of the source analog.

3. Jaccard Coefficient (*JCoef*): To measure the mapping in relation to both source and target analogs we use the Jaccard coefficient (Jaccard, 1901), which is a measure of the similarity between two finite sets. In this case, the mapping is treated as the intersection between the source and target graphs (Abgaz, et al., 2016-b). This measure incorporates the size of the mapping together with the source and target graphs. It produces a value between 0 and 1 where values near 1 indicate greater structural similarity (homomorphic).

A Spearman rank order correlation indicated the Jaccard coefficient did not identify creative analogies r_s=0.078 (*p*=0.41). Another thought provoking result as the proportion of the two analogs that participate in a comparison appears to have no impact on the resulting creativity – bearing in mind the small sample size involved in this paper.

So, metrics that focus purely on the topology of the analogs do not appear to be adequate in identifying creative comparisons. Of course, we need to remain cognizant that this does not mean the mapping is irrelevant to creativity, as this evaluation focused only on the best analogies identified for each target. But perhaps semantic factors will prove more fruitful in our quest for creative analogies.

Generic Space Metrics

Dr Inventor quantifies the semantic similarity between mapped counterparts, which are aligned by the mapping process. The generic space generalizes across each pair of mapped items from the two publications (for structural reasons, these may not necessarily be the most semantically similar pairings). Dr Inventor pays particular attention to mapped but not identical items as semantically distant "between domains" comparisons are often seen as the hallmark of creative comparisons (Koestler, 1964).

In this paper, we estimate semantic similarity using the *Lin* (Lin, 1998) similarity metric, which in turn is based on the WordNet lexical database. The Lin metric produces results in the range [0-1] with values closer to 1 indicating a greater degree of semantic similarity.

Dr Inventor maintains independent metrics for noun and verb based similarity for several reasons. Firstly, WordNet's verb-based entailment hierarchy is generally shallower than its noun hierarchy and this has implications for the *Lin* metric that incorporates the "lowest common subsumer" in its calculations. This causes potential problems and inconsistencies when comparing the relative influence of noun-based and verb-based differences in comparisons. Secondly, a mapped predicate typically involves two mapped nouns but only one mapped verb, which might easily lead noun based similarity to overwhelm the relational similarity. Finally, structure mapping theory (Gentner, 1983) can help differentiate between "literal similarity" and "analogy" by favoring mappings between unrelated nouns but similar verbs.

1. Conceptual Similarity (*ConSim*): Conceptual Similarity measures the similarity of paired concepts (nouns) using the Lin metrics. For example, a boat and a car, *Lin(boat#n, car#n)= 0.7198* , share more commonality than a boat and a "cat", *Lin(boat#n, cat#n)= 0.1647.* Our mapping algorithm selects the pair with a higher similarity score when it is presented with such a choice. For a given analogy, we use the mean value between the nouns involved in the mapping.

A Spearman rank-order correlation coefficient revealed a moderate negative relationship between user ratings and the estimate of conceptual similarity (*ConSim*) r_s= -0.442, *p*=0.09. This was an interesting finding as analogies involve, almost by definition, comparisons between different objects. So, a low (though positive) correlations was quite expected, but this strong negative correlation was quite surprising – despite not quite reaching a level of statistical significance. This finding suggests that creative comparisons don't just pair objects with little similarity to one another - but involve objects that are notably and quantifiably *dis*-similar to one another! This, finding was even more surprising given the semantic homogeneity associated with using papers only from SIGGRAPH.

This finding can be seen as allied to the idea that creative comparisons arise from "between domains" comparisons typically involving dissimilar objects (Blanchette & Dunbar, 2000). This conceptual dissimilarity can also be seen as comparable to factors such as the "tension" associated with between domains comparisons.

2. Relational Similarity (*RelSim*): Relational Similarity measures the similarity of paired relations (verbs) again using the Lin metric, ensuring that lexically ambiguous terms are interpreted in their verb sense only. To estimate relational similarity, we use the mean relational similarity, averaged across all mapped relations in the comparison.

A Spearman rank-order correlation coefficient revealed a moderate relationship between user ratings and the estimate of relational similarity (*RelSim*) between analogous publication r_s=0.430, *p*=0.10. This positive but weak relationship was surprising because we expected it to be even stronger. Relational similarity can be seen as the semantic foundations of systematicity theory (Gentner, 1983).

3. Latent Semantic Analysis (*LSA*): Previous studies have shown that similarity as estimated using LSA is not useful in identifying detailed analogies (Ramscar & Yarlett, 2003) or creative comparisons (Abgaz, et al., 2016-b). A Spearman Rank-order correlation between the average creativity rating and the LSA for each comparison was r_s = -0.6201 (p<0.05), suggesting that increasing the semantic distance between analogs had a positive impact on users' creativity. Our results show that conceptual and relational similarity have very different influences on users' perceptions of creativity.

Inference Metrics (Blended Space)

We next present several metrics related to the inferences and creation of the blended space. Dr Inventor counts the inferences it generates from each comparison, allowing identification of the more creative, if not profligate comparisons.

1. Number of Inferences (*NumInfs*): Finally, we examined the impact of the number of inferences upon the average rating awarded to a comparison. The number of grounded inferences indicates the amount of new information the analogy provides. A Spearman rank order correlation r_s=0.286 (p=0.21) did not show any reliable influence from the number of inferences on users' creativity.

While the Spearman correlation tests for the presence of linear associations between variables, the Wilcoxon signed rank test looks for differences between two populations. A Wilcoxon paired signed rank test between the number of inferences and the user ratings for each analogy revealed that the null hypothesis could not be rejected (V=7, p<0.05). Additionally, a Pearson Product Moment correlation of 0.613 was identified. Thus, we infer that the number of inferences had an impact on the creativity ratings awarded by Dr Inventor's users.

2. Novelty of Inferences (*ObservedNovelty*): The raw count of the inferences doesn't address the properties of *novelty* and *quality* that are central to creativity (Boden M. A., 2004). We situate our estimation of novelty within the socially creative context (Corneli, 2016) of recent publications in this conference series. Thus, we assess novelty using an *n*-gram approach (Abgaz, O'Donoghue, Smorodinnikov, & Hurley, 2016) derived from the SIGGRAPH corpus, with the resulting *n*-grams estimating the novelty of inferences.

Firstly, a tri-gram model was constructed from the entire SIGGRAPH corpus of 721,301 triples, with 604,873 distinct triples (ignoring duplicates). For example, a common inference was found to be: *we (introduce, algorithm),* however most inferences were novel with respect to these tri-grams.

The corresponding bi-gram model was constructed allowing "piecemeal" evaluation of the (relative) novelty of unfamiliar inferences. So, the *Subj-Verb, Verb-Object* and *Subject-Object* combinations can be evaluated in a piecemeal fashion, allowing Dr Inventor to compare the degree of novelty contained within a novel inference. Lower bigram probabilities arise from truly novel combinations of information, indicating a greater level of creativity. Let i signify an individual inference and $|i|$ represent its trigram frequency in the repository, then the *novelty* of the inference, $N(i)$ is given by

$$N(i) = \begin{cases} 0 & |i| > 0 \\ 1 - P(i) & |i| = 0 \end{cases}$$

where $P(i)$ is the bi-gram probability.

For a given analogy producing m individual inferences, the novelty score is calculated as the average novelty scores of the individual inferences.

The Spearman rank order correlation between the observed novelty score and the user ratings r_s=0.42 (p=0.11). While this result was not reliable, partly due to the small sample size, it does suggest that novelty of inferences is a factor in creativity ratings. Thus, we argue that this shows that the novelty of inferences might be a factor influencing users' perceptions of creativity. Further evaluations will be required to explore this factor in greater detail.

A Research Hypothesis by Dr Inventor, by N.C.C.A., Bournemouth University, UK

A research hypothesis created by Dr Inventor's led to the following research project. '*Curve-Skeleton Extraction from Incomplete Point Cloud* (2009)' describes an algorithm for curve skeleton extraction from point clouds, where large portions of data are missing during 3D laser scan. Dr Inventor system has identified '*Fast Bilateral Filtering for the Display of High-Dynamic-Range Images* (2002)' as analogous, presenting a technique to display high-dynamic-range images, which reduces the contrast while preserving details.

The **creative analogy** sees both papers focus on the reconstruction of hidden structural information. Paper-1 solves a 3D problem of incomplete vertex data containing **holes** (caused by self-occlusions during 3D laser scanning). Paper-2 solves a 2D problem in images with poor light management, with under-exposed and over-exposed **areas**, and light behind the main character.

Proposed solution: The papers are from different Computer graphics domains (Modelling and Image processing) and their methods cannot interpolate with each other. Interestingly, "holes" in the problem are mapped with "areas" in the source paper. Analogous examples from existing work has lead us to the following developments.

Inspired analogy: We have explored new ideas through Dr Inventor to learn; How to rebuild and animate 3D models automatically and reconstruct hidden structure more efficiently. A new idea has been generated after a case-study was carried out from the literature of 18 papers by Dr Inventor. We have found a new method to represent natural flower shape, flower blooming and the decay process, using an Ordinary Differential Equation (ODE)-based surface modelling & simulation technique. Interestingly, the analogy paired "area" with "hole", normally seen as opposites.

Shape representation of flower is challenging and interesting topic which has attracted the many researchers. The shape of flower consists of a multi-layer architecture (petals, stigma, and stems). Each part of a flower involves a complex geometrical deformation such as bend, stretch, shrink and curl. Various techniques (Data-driven, Sketch-based, Point-based and Image-based) are popular, but face challenges such as the geometry of high fidelity and missing-captured data.

Advantages of our new method: In order to address the existing challenges for the shape representation and simulation of flower, we present a single framework which uses ODE-based surface modelling & simulation technique to solve geometry structural information more efficiently. Our method is very useful for 3D modelling and simulation that creates realistic flower shapes with a small data size.

3. Novelty Relative to Other Inference (*PredictedNovelty*): We also compared inferences against all *other inferences* generated from all possible analogies from our corpus. Dr Inventor explored over 1.3 million analogies producing 225,230 inferences, of which 151,200 were unique (ignoring duplicates). We might think of this collection as inferences likely to arise (at least analogically) from the collected wisdom contained in the corpus. Tri-gram and bi-gram models were used to estimate the novelty of inferences in relation to all other inferences. However, a Spearman rank order correlation between the novelty of the SIGGRAPH inferences and the user ratings was r_s=0.048 (p=0.446) showed that this was not an influencing users' perceptions of creativity.

Multi-Space Metrics

1. Analogical Similarity (*AnaSim*): This evaluates the mapping in terms of structural and semantic factors, combining Jaccard's coefficient with conceptual and relational similarity:

$$AnaSim = \left((RelSim + ConSim)/2\right) * JCoef$$

A Spearman Rank-order correlation between the average creativity rating and *AnaSim* showed r_s=0.349 (p=0.16). Again, this is suggestive of a mild relationship between *AnaSim* and creativity ratings.

2. Overall Similarity Indicator (*OverallSim)*: Finally, we look at a theoretically driven combination of these metrics giving a single usable means of selecting the most creative comparisons. Rather than using the number of metrics we employ an exponential squashing function scaling the number of inferences to the range [0…1] in order to select analogies with a moderate number of inferences, while comparisons offering huge number of inferences will gain little advantage. This intervention was made to avoid overwhelming users with too many inferences.

We combine overall analogical similarity (*AnaSim*) score with the scaled inference metric. Novelty is used to determine whether we should include the inference during the presentation to the users. This metric has been used to order the inferences, allowing users to focus on the more informative ones first, when many are available.

$$OverallSim = AnaSim * \text{numInf}$$

Final Evaluation

The influence of individual metrics may differ when used in combination with one another, so we explore linear combinations of these different facets using principal component analysis (PCA) (Jolliffe, 2002). PCA is often used to explain the variance within data and while a few of the metrics above (e.g. *AnaSim*) are already linear combinations of more primitive metrics, this evaluation focuses exclusively on the primitive metrics. A PCA analysis was conducted and results show that the first principal component accounting for 63% of the variance was formed the following combination of factors:

-0.533ConSim+0.529RelSim+0.485numInf+0.311PredictedNovelty+0.289ObservedNovelty+0.144JCoef

The first (and thus largest) principal component indicates (by dint of the -0.533) that more creative comparisons involve conceptual (noun-based) *dis*-similarities, as discussed earlier. Conversely, relational (verb-based) similarity appears to be a factor in creative comparisons, as are greater numbers of inferences. The two novelty scores are also important factors in this principal component, helping remove analogies suggesting uncreative inferences. Four principal components account for all variance in this collection.

This combination of factors may allow identification of better and even more creative comparisons in future versions of the Dr Inventor creativity enhancement tool.

Conclusions

This paper presents a computational system called Dr Inventor that explores novel analogical comparisons between published research documents. Dr Inventor combines a lexical analysis phase with a model of analogical thinking, which forms the core of our model of analogical reasoning and conceptual blending. This paper focuses on the specific problem of identifying the qualities of creative comparisons that help make them creative. Qualities and associated metrics of comparisons were explored, derived from the mapping (counterpart projection), generic space and blended spaces, focusing on both semantic and topological factors.

The main finding in this paper concerns the characteristics of analogy-based comparisons and their potential use as predictors of creative comparisons. Results suggest that it is not the strength of the analogy (or counterpart projection), but rather it is the inferences and their novelty that are the hallmarks of creativity. In particular inferences and their novelty play a significant factor in the creativity ratings given by expert users of the Dr Inventor system.

Dr Inventor treats all inferences as "additive" to the existing body of knowledge, but adding the ability to detect contradictory beliefs might require an incompatible/alternative "belief space". Dr Inventor might thereby well be extended to support novel and alternative belief spaces of Boden's Transformational Creativity. The main challenge lies in determining greater semantic specifity for its conceptual and relational nodes, requiring advances in language processing, semantic tagging and ontology.

We see Dr Inventor's model of non-literal similarity as being one possible approach to supporting creative reasoning across research disciplines. While Dr Inventor has currently only been tested on documents from SIGGRAPH, we believe it points the way for useful progress. We believe that systems like Dr Inventor may offer vital leverage in promoting inter-disciplinary thinking and research.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement no 611383.

References

Abgaz, Y., O'Donoghue, D. P., Smorodinnikov, D., & Hurley, D. (2016). Evaluation of Analogical Inferences Formed from Automatically Generated Representations of Scientific

Publications. Artificial Intelligence and Cognitive Science (AICS). Ireland.

Abgaz, Y., O'Donoghue, D. P., Hurley, D., Saggion, H., Ronzano, F., & Smorodinnikov, D. (2016-b). Embedding a Creativity Support Tool within Computer. ECAI - Modelling & Reasoning in Context.

Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In Handbook of research on assessment technologies, methods, and applications in higher education (pp. 65-77). IGI Global.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. Memory & Cognition, 28(1), 108-124.

Boden, M. A. (1996). Dimension of Creativity (2nd ed.). Cambidge, Massachusetts: MIT Press.

Boden, M. A. (2004). The creative mind: Myths and mechanisms. Psychology Press.

Bruza, P., & Weeber, M. (2008). Literature-based discovery. Springer.

Cordella, L., Foggia, P., Sansone, C., & Vento, M. (2004). A (sub)graph isomorphism algorithm for matching large graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26, 1367-1372.

Corneli, J. (2016). An institutional approach to computational social creativity. Intl. Conf. on Computational Creativity (ICCC), (pp. 131-138).

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. Cognitive science, 22(2), 133-187.

Gentner. (1983). Structure- mapping: A theoretical framework for analogy. Cognitive science, 7(2), 155-170.

Gentner, D., & Smith, L. (2012). Analogical reasoning. In V. S. (Ed.), Encyclopedia of Human Behavior (2nd Ed.). (pp. 130-136). Oxford: Elsevier.

Goel, A. K., Creeden, B., Kumble, M., Salunke, S., Shetty, A., & Wiltgen, B. (2015). Using Watson for enhancing human-computer co-creativity. AAAI Fall Symposium.

Hurley, D., Abgaz, Y., Ali, H., & O'Donoghue, D. P. (2016). Expert and Corpus-Based Evaluation of a 3-Space Model of Conceptual Blending. ECAI - Evaluatring General Purpose AI.

Jaccard, P. (1901). Distribution flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 241-272.

Jolliffe, I. (2002). Principal Component Analysis (second edition ed.). New York: Springer-Verlag New York, Inc.

Jordanous, A. (2012). A standardised procedure for evaluating creative systems. Cognitive Computation, 4(3), 246-279.

Juršič, M., Bojan, C., Tanja, T., & Lavrač, N. (2012). Cross-domain literature mining: Finding bridging concepts with CrossBee. Proc. 3rd Intl. Conf. on Computational Creativity ICCC, (pp. 33-40).

Kantosalo, A., & Toivonen, H. (2016). Modes for Creative Human-Computer Collaboration: Alternating and Task-Divided Co-Creativity. Intl. Conf. Computational Creativity ICCC.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. Review of General Psychology, 13(1), 1-12.

Koestler, A. (1964). The Act of Creation.

Krippendorff, K. (2011). Agreement and Information in the Reliability of Coding. Communication Methods and Measures, 5(2), 93-112.

Lin, D. (1998). An Information-Theoretic Definition of Similarity. San Francisco, CA, USA: Morgan Kaufmann.

Martins, P., Pollak, S., Urbancic, T., & Cardoso, A. (2016). Optimality Principles in Computational Approaches to Conceptual Blending: Do We Need Them (at) All? 7th International Conference on Computational Creativity.

O'Donoghue, D. P., & Keane, M. T. (2012). A Creative Analogy Machine: Results and Challenges. Proc Intl. Conf. on Computational Creativity (ICCC), (pp. 17-24).

O'Donoghue, D. P., Abgaz, Y., Hurley, D., & Ronzano, F. (2015). Stimulating and simulating creativity with Dr inventor. Proc. Int. Conf. on Computational Creativity ICCC, (pp. 220-227).

Pinel, F., & Varshney, L. R. (2014). Computational Creativity for Culinary Recipes. In CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14) (pp. 439-442). New York, NY, USA: ACM.

Ramscar, M., & Yarlett, D. (2003). Semantic grounding in models of analogy. Cognitive Science, 1, 41-71.

Ronzano, F., & Saggion, H. (2015). Dr. Inventor Framework: Extracting Structured Information from Scientific Publications. In Discovery Science (pp. 209-220). Springer.

Schorlemmer, M., Smaill, A., Kühnberger, K., Kutz, O., Colton, S., Cambouropoulos, E., & Pease, A. (2014). Coinvent: Towards a computational concept invention theory. Intl. Conf. Computational Creativity.

Spangler, S., Wilkins, A. D., Bachman, B. J., Nagarajan, M., Dayaram, T., Haas, P., & Regenbogen, S. (2014). Automated hypothesis generation based on mining scientific literature. Proc. 20th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, (pp. 1877-1886). 2014.

Toivonen, H., & Gross, O. (2015). Data mining and machine learning in computational creativity. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(6), 265-275.

Van Mieghem, P. (2013). Can the topology of two graphs be compared by one number? Delft University of Technology.

Veale, T., & Keane, M. T. (1997). The competence of sub-optimal theories of structure mapping on hard analogies. IJCAI, (pp. 232-237.).