Evaluation of Analogical Inferences Formed from Automatically Generated Representations of Scientific Publications

Yalemisew Abgaz, Diarmuid O'Donoghue, Dmitry Smorodinnikov, Donny Hurley

Department of Computer Science, Maynooth University, Maynooth-Ireland
{abgaz.yalemisew,Diarmuid.ODonoghue,Donny.Hurley}@nuim.ie

Abstract. Humans regularly exploit analogical reasoning to generate potentially novel and useful inferences. We outline the Dr Inventor model that identifies analogies between research publications, describing recent work to evaluate the inferences that are generated by the system. Its inferences, in the form of *subject-verb-object* triples, can involve arbitrary combinations of source and target information. We evaluate three approaches to assess the quality of inferences. Firstly, we explore an *n*-gram based approach (derived from the Dr Inventor corpus). Secondly, we use ConceptNet as a basis for evaluating inferences. Finally, we explore the use of Watson Concept Insights (WCI) to support our inference evaluation process. Dealing with novel inferences arising from an ever growing corpus is a central concern throughout.

1 Introduction

An analogy is a comparison between two concepts (the source and target), where the comparison itself is somewhat novel and interesting due to differences between the two concepts. Based on their perceived similarities and subsequently extending them is called analogical inference and such inferences often cast new information onto the target using information obtained from the source. Such comparisons aid our understanding of less well-known concepts, by "re-cycling" other information. Analogy requires systematic comparison of the structure of the two concepts involved. Analogical reasoning is used in education [**1**], scientific discovery [**2**], and to explain and discover new knowledge about less-known systems. However, analogical inferences are not be always true and can be misleading [**3**].

Analogical reasoning [**4**] focuses on three main processes: 1) Retrieval of a *source* for a given *target*, 2) Mapping [**5**] [**6**] the source to the target by structural alignment and inferences generation [**2**], 3) Evaluation, where inferences are judged [**3**] and potentially rejected. Elsewhere we [**7**] [**8**] described our analogy model ("*Dr Inventor*") that discovers analogies between scientific documents– but validating the resulting inferences is crucial to the successful use of Dr Inventor. This paper describes an inference evaluation model for use by Dr Inventor that aims to remove invalid inferences while preserving the good inferences. Thus, we present an *n*-gram based familiarity analysis method and try to answer the following main questions: 1) How to

differentiate familiar/good inferences from those that are unfamiliar/bad inferences, 2) How different knowledge sources can be used and how they affect analysis of familiarity of inferences, and 3) which metrics can be used and how can they be tuned to measure the degree of familiarity of the inferences. We expect to find similar inferences: 1) made by other papers, 2) exhibit strong associations between the *subjects*, *verbs* and *objects*, and 3) be familiar to human evaluators.

The rest of this paper is organized as follows. Section 2 focuses on some related work in the area of analogical reasoning and evaluation of inferences and Section 3 gives highlights for our analogy mapping model with a well-known analogy example and how inferences are generated, followed by a detailed examination of our validation model. In Section 4 we present the experiment and evaluation results. Finally, section 5 focuses on conclusion and future work.

2 Related Work

Thinking with analogies is a form of structure driven reasoning that appears to play a role in many different areas of intelligence. Computational modelling of this cognitive process is enabled through Gentner's [**2**] Structure Mapping Theory (SMT). This theory posits that to find the analogical similarity between the source and target, we must identify the largest common sub-graph between the source and target structures. Since its inception, SMT has led to focused work on distinct phases of analogy, particularly on the retrieval and mapping phases.

The key algorithm for generating inferences is called CWSG - Copy With Substitution and Generation [**9**]. Building upon the inter-graph mapping, CWSG identifies structures from the source that can be transferred to the target. But CWSG is blind to the potential credibility of its inferences. As noted by [**10**] and others, analogy is a profligate inference mechanism giving rise to our inference evaluation system.

Several attempts have been made to evaluate analogical inferences. [**3**] argue that the strength of analogical inferences depends on the level of similarity between a source and a target. Humans are highly selective analogizers and they focus on relational pattern completion which (they argue) effectively filters out bad inferences. Dr Inventor does not have access to the expertise required to support such filtering, so we shall adopt a different approach. [**11**], used analogical reasoning techniques (inference rules) to generate the facts, it uses humans to evaluate the plausibility of the inferences and 35.6% of the inferences express new true inferences. However, it lacks any automatic evaluation method and thus, is not applicable to Dr Inventor.

3 Inferences with Dr Inventor

3.1 Generating Graph Representations of Research Information

This section outlines the preprocessing and mapping phases of Dr Inventor. It accepts academic publications as input (such as PDF documents). Text extraction using PDFX resolves complications like: headers, footers, equations, table, page numbers etc.

Identified text is passed to a state-of-the-art natural language processing pipeline to generate dependency trees. The parser includes a classifier that classifies sentences according to their rhetorical category (abstract, background, approach, outcome, future work). Details of the text processing pipeline are discussed in [**12**].

Using the output of the text processing pipeline, we convert the information from the dependency tree to a Research Object Skeleton (ROS) graph that efficiently represent the concepts (nouns) and relationships (verbs) of each sentence. A ROS-graph captures contents of the input text in a form of subject-verb-object triples constructed from each sentence. Using co-reference resolution that is built into the dependency parser, multiple occurrences of the same concept are uniquely represented within each ROS. Co-reference resolution greatly improves ROS graph quality, linking words like "it" to their referents. Each ROS represents each concept uniquely across the entire document. Interestingly, this echoes a recent work on embodied cognition identifying three reasons for unique representation [**13**].

All triples for a document form a large interconnected ROS graph, though some disconnected triples also occasionally arise. Subgraphs can be extracted for lexical or rhetorical subsection of papers. We demonstrate ROS generation with the following example. The abstract of the source paper ("*Gaussian KD-tree for fast high-dimensional filtering*"[1]) is found analogous to the target paper ("*Linear Combination of transformation*"[2]) and the method applied to the source is analogically applicable to the target paper's problem. The two abstracts are transferred into a ROS graph (**Fig. 1**.)

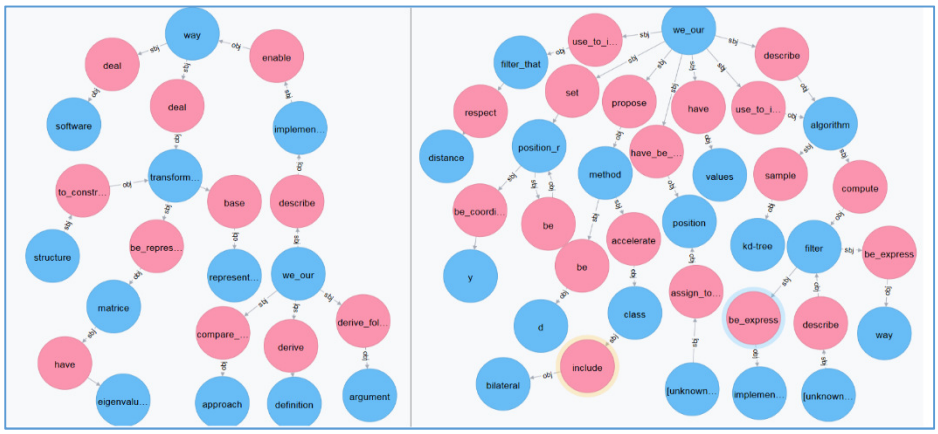

Fig. 1. A snippet of ROS Graphs for Target (Left) and Source (Right) Paper.

Source Paper. *We propose a method for accelerating a broad class of non-linear filters that includes the bilateral, non-local means, and other related filters. These filters can all be expressed in a similar way: First, assign each value to be filtered a position in some vector space. Then, replace every value with a weighted linear combination of all values…*

[1] http://dl.acm.org/citation.cfm?doid=1576246.1531327
[2] http://dl.acm.org/citation.cfm?doid=566654.566592

Target Paper. *Geometric transformations are most commonly represented as square matrices in computer graphics. Following simple geometric arguments we derive a natural and geometrically meaningful definition of scalar multiples and a commutative addition of transformations based on the matrix representation, given that the matrices have…*

3.2 Analogy Mapping

After constructing the ROS graph, Dr Inventor commences the analogical mapping process, which is based on structure mapping theory [**2**] [**14**]. It uses subgraph isomorphism for finding the best alignment between the source and target graphs.

A ROS is a form of attributed relational graph with labels as types to identify the conceptual category of the graph as "noun" or "verb", the mapping process only maps nodes that are in the same conceptual category. This constraint further reduces the time required by the graph matching process by significantly reducing the search space. Our algorithm ranks nodes based on some centrality metrics (Degree, Node rank) and starts the mapping from the most "central" node, to further expedite this process. The graph matching is primarily guided by structure (comparing in degree, our degree etc) and complemented by the WordNet [**15**] based Lin semantic similarity metric [**16**] when a single node of the target structurally matches two or more candidate nodes from the source. Thus, mapping occurs between two most structurally similar nodes and when two or more nodes have equal importance, we select pairs that have highest semantic similarity. We customized the (sub-)graph isomorphism algorithm (VF2) [**17**] to identify analogies between two ROS graphs. VF2 was selected due to its efficiency in search time, as Dr Inventor is expected to explore many mappings in order to find novel and useful comparisons. A snippet of the mapping of the example is given in **Table 1**.

Table 1. Sample Mappings between abstracts of the two papers

Source	Target	Label	Sim Score	Source	Target	Label	Sim Score
Position	Software	Noun	0.261	Class	Definition	Noun	0.096
Accelerate	Drive	Verb	0.553	d	Argument	Noun	0.000

3.3 Analogical Inference

In this section we will discuss our proposed analogical transfer and evaluation system. Inference generation uses the mapping pairs, the source ROS and the target ROS to identify the candidate inferences. There are constraints we defined to identify candidate nodes for transfer – or candidate inferences.

The Dr Inventor system is designed as a creativity support tool, identifying novel comparisons between publications. This novelty requirement inevitability results in inferences that involve new (previously unseen) inference that must be evaluated for their likely usefulness. Novel inferences involve novel combinations of *subject, verb* and *object* terms originating in the two publications. A typical novel inference might involve two terms from the source publication (say) while the other is from the target – as exemplified below.

Source: $subject_s$, $verb_s$, $object_s$.
Target: $subject_t$, $verb_t$, $object_t$.
Candidate Inference: $subject_t$, $verb_s$, $object_t$.

Some of the bad triples include "mirror ask butter" and "bridge phone sun". Particular challenges for Dr Inventor include: 1) evaluating novel combinations of subject, verb and object 2) evaluating candidate inferences between arbitrary pairs of publications 3) dealing with an ever changing corpus of documents.

All candidate inferences may be referred back to mapped pairs – enabling use of the "grounded inference" constraint. This means, for a node to be considered as candidate for the transfer, it should be linked to one or more of the nodes that are already mapped to the target node. We split the general constraint into three simple constraints: constraint on verbs, constraint on nouns and constraint on edges (Fig 2).

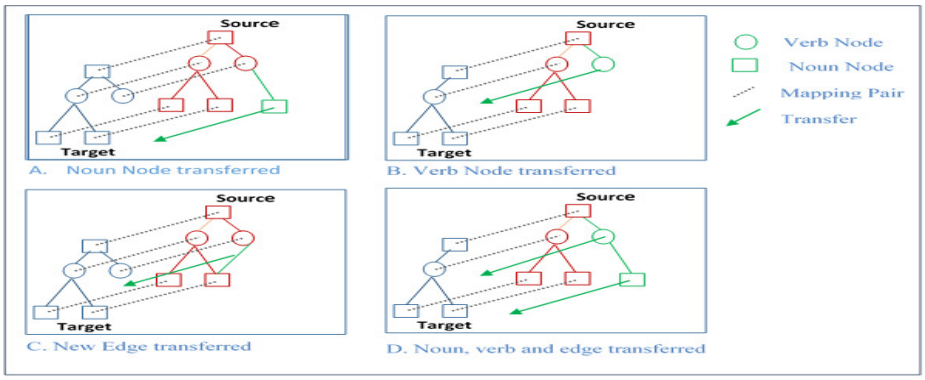

Fig. 2. Analogical transfer from source to target

Candidate verb inference. A verb node is considered as a candidate for inference if there is an edge that links it to the source end of a mapping pair or if it forms a link between two transferred noun nodes (Figure 2. B or D). ***Candidate noun inference.*** A noun node is considered as a candidate for inference if there is an edge that links it to the source end of a mapping pair or if it forms a link between two transferred verb nodes (Figure 2. A or D). ***Candidate edge inference***. An edge is transferred as a candidate inference if it is linked to the candidate verb node or the candidate noun node or if it links two mapped nodes in the source (Figure 2. C). Only inferences that match these constraints are considered a viable inferences, being sufficiently connected to the underlying mapping. It further requires to put additional constraints to determine the number of nodes that should be transferred. If the number of nodes transferred to the target is greater than the number of mapping nodes, it becomes less usable. One example of an inferred triple from the previous analogy is *"definition include bilateral"* where "*include bilateral*" is transferred and attached to "*definition*".

Familiarity as a Basis for Validating Inferences. The two defining characteristics of creativity are Novelty and Quality and in this paper we explore the use of "familiarity" as a basis for the joint evaluation of novelty and quality. We start with an

n-gram-based technique that deals well with familiar inferences, followed by partial evaluation of subj-verb, verb-obj or subj-obj pairs as partial evaluation of inferences. The n-gram approach is extended by exploring several "smoothing" techniques to estimate the familiarity of unseen triples. Finally, we extend these techniques by exploring ConceptNet and Watson Concept Insights for assessing quality of novel inferences. While these approaches estimate quality, it should be pointed out that for any "collection" of inference to be considered truly creative, we expect that a number of these inferences will not successfully validated. Any collection of inferences all of which are familiar can be rejected as it does not offer sufficient novelty! Conversely, if all inferred information is invalidated (thus is considered very novel) this too could be rejected as a useful comparison as it could place too high burden on the user.

Our concern with creative comparison and creative inferences is that the resulting creativity should contain an appropriate level of novelty. Work is currently ongoing to assess the optimum balance of familiar and novel information with which to serving users creative needs.

3.4 Validation of Inferences with *n*-grams.

Inference validation focuses on evaluating the degree of strength of a triple using familiarity analysis. We use n-gram based methods to calculate the familiarity of inferences. We later used these scores to rank triples based on their familiarity, showing how they can be applied to the novel triples that Dr Inventor aims to generate.

n-gram model. We employ an *n*-gram model to evaluate how good a given sequence of words fit together. Thus the probability of a series of words is given by

$$p(w_1, w_2, w_3 \dots w_n) = p(w_1)p(w_2|w_1)p(w_3|w_1, w_2) \dots p(w_n|w_1 \dots w_{n-1}). \quad (1)$$

This formula can be simplified by applying the Markov Assumption, which states that the probability of a word in a text depends only on *n-1* preceding words. In our case the sequence of words is in a form of *"subject-verb-object"*. As for this particular work, unigrams, bigrams and trigrams are of prime interest. The probability of a word depends only on one preceding word in bigram model and on two preceding words in trigram model. For unigrams, a probability of a word is independent of the preceding words. To estimate $p(w_i|w_{i-1})$ we need two components: 1) the count of the bigram $w_{i-1}w_i$ and 2) the count of all possible bigrams where w_{i-1} is the first word.

Now we explore the *n*-gram models as our subject-verb-object inferences fit the *n*-gram models. The unigram approach takes all the individual elements of the triple and calculates the probability independent of the other remaining elements. But this approach gives us little information as it doesn't tell us any information on how well each of the terms "fit together". The bigram approach calculates the probability of one element in relation to the other two elements (in the form of two separate bigram probabilities). Thus, the probability of a triple is given by

$$p(s, v, o) = p(s| < \text{start} >)p(v|s)p(o|v)p(< \text{end} > |o). \quad (2)$$

<start> indicate beginning and <end> indicate the end of a triple. And trigrams are calculated as

$$p(s, v, o) = p(s| < \text{start} >)p(v|s < \text{start} >)p(o|sv)p(< \text{end} > |vo), \quad (3)$$

where, s is Subject, v is Verb, o is Object.

Using a trigram approach, for example, we can calculate the probability of "we-describe-algorithm" as p("we", "describe", "algorithm") = p("we") p("describe" | "algorithm") p("algorithm" | "we", "describe"). Such n-gram model will allow us to calculate the probability of one word occurring with another in such sequence.

However, the n-gram model has an inherent problem in that if any of the probabilities are zero, then the whole probability become zero. This makes the familiarity analysis useless. To avoid this problem different methods are proposed. First, we apply synonym substitution method and then we consider two smoothing approaches called additive smoothing and Good-Turing smoothing.

Additive Smoothing. We explore additive smoothing to avoid the zero probability by replacing r occurrences of n-gram in a corpus with $r + \delta$ occurrences. δ needs to be a small number between 0 and 1. This changes the probability to

$$p_{\text{add}}\left(w_i \middle| w_{i-n+1}^{i-1}\right) = \frac{\delta + C(w_{i-n+1}^{i})}{\delta|V| + \sum_{w_i} C(w_{i-n+1}^{i})}, \quad (4)$$

where V is a set of all words considered c is the count of the corresponding word.

Good-Turing Smoothing. Good-Turing smoothing uses the count of events we have seen once to predict the count of things we have never seen. This strategy tried to estimate the weight of the unseen events by reducing the probability mass of already observed events. We introduce a notation N_r a frequency of frequencies, meaning how many things occurred with frequency. Let's assume that some n-gram occurs r times in our database. According to classical Good-Turing, should be replaced by r*, where

$$r^* = (r + 1)\frac{N_{r+1}}{N_r}. \quad (5)$$

Then the probability of an n-gram $\boldsymbol{x}$ is calculated as

$$p(\boldsymbol{x}) = \frac{r^*}{|V|} \quad (6)$$

Evaluation with ConceptNet[3]: ConceptNet (v5.4) is a database of concepts and their inter-relationships, representing common sense background information. Interestingly, ConceptNet provides a numeric measure to estimate degree of association between concepts. In the following sections we use it to evaluate the strength of inferred triples.

[3] http://conceptnet-api-1.media.mit.edu/

Evaluation with Watson Concept Insights: Watson Concept Insights (WCI) provides an API [**18**] that computes the strength of conceptual associations, which we use to evaluate inferences. The concept graph used by the WCI service has been derived from the English language Wikipedia. We also use WCI as another source of formalized knowledge to evaluate individual inferences. WCI is selected particularly for its fine grained confidence score.

4 Experiments and Results

For the experiment we generated different collections of "subject-verb-object" triples as data sets from three different sources. Then we evaluated these dataset against their respective knowledge sources. Finally, we included human evaluation of the datasets and compare them with the results we get from the system. Dr Inventor dataset contains 572,496 triples extracted from 957 computer graphics papers published between 2001 and 2015 from SIGGraph and SIGGraph-asia following the procedure in section 3.1.

4.1 Overview of Evaluation Procedure

Ten human evaluators were recruited to evaluate inferences, all being selected from the computer science discipline and include lecturers, post-doctoral researchers and postgraduate students. The respondents were given the triples in a random order and the evaluation was separated in to two parts.

First raters evaluated the domain specific triples (computer graphics) from Dr Inventor corpus by randomly selecting 1000 triples from the Dr Inventor collection. Their familiarity scores were calculated using both Additive and Good-Turing smoothing methods. Then we took 20 *good* inferences from each method (40 triples together) and 20 *bad* inferences (another 40 triples) and give them the 10 evaluators. The expert evaluators rated the triples on a scale of 0 to 5, where 0 denotes unfamiliar, 2-3 medium familiarity and 5 represents high familiarity.

Second, raters evaluated domain independent triples. This evaluation used Random Lists[4] to generate random English nouns and verbs to form (generally) bad triples and we used the corpus of contemporary American English (COCA) to identify good triples. Since COCA contains sentences extracted from fiction, popular magazines, newspapers and academic texts, we verified that the triples extracted from this corpus were meaningful and familiar. We extracted 29 familiar triples and 31 bad triples and by combining nouns and verbs randomly. Some of the bad triples include *"mirror ask butter"* and *"bridge phone sun"*.

4.2 Results and Discussion

Evaluation results from Dr Inventor Triples. Table 2. shows the threshold values that were determined for 1000 evaluated triples. The score is computed using the

[4] https://www.randomlists.com/

probability distribution of N-grams and threshold is decided based on the distribution of the familiarity score over a large collection.

Table 2. Thrshold values for familiarity of Dr inventor triples.

Score	Additive smoothing	Good-Turing smoothing
Low	$0 < score \leq 1.67 \times 10^{-14}$	$score \leq 9.91 \times 10^{-10}$
Medium	$4.99 \times 10^{-10} < score < 1.67 \times 10^{-14}$	$2.28 \times 10^{-5} < score < 9.91 \times 10^{-10}$
High	$score \leq 4.99 \times 10^{-10}$	$score \leq 4.99 \times 10^{-10}$

Using the above score interpretation, 70.8% of the triples are assigned the same rating category by both methods. 125 scored "high", 455 score "medium" and 128 scored "low". So, we have 70.8% agreement between the two methods. We evaluated the resulting 40 high score triples and 40 low score triples by experts, to investigate how close our approach is to human evaluators.

Table 3. Human evaluation results of triples rated as high (left) and low (right) by our system

Smoothing technique	Score	Triples evaluated as "high"		Triples evaluated as "Low"	
		No	Percentage	No	Percentage
Additive	High	17	85%	3	15%
	Medium	2	10%	12	60%
	Low	1	5%	5	25%
Good-Turing	High	15	75%	3	15%
	Medium	4	20%	8	40%
	Low	1	5%	9	45%

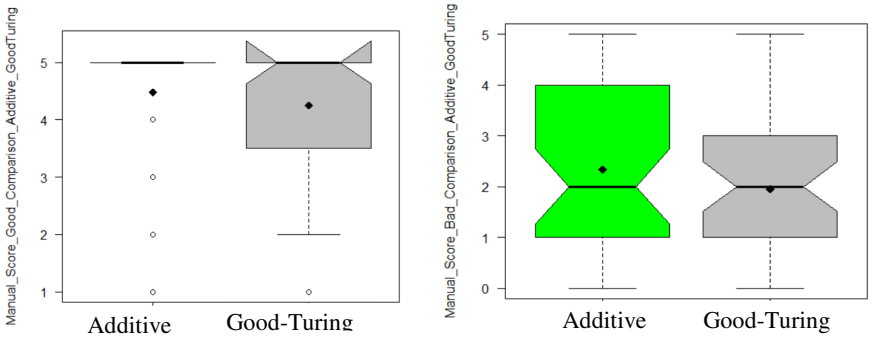

Fig. 3. Additive and Good-Turing comparison for "high" (left) and "Low" (right).

Table 3 shows a comparison between our proposed methods and human evaluation. Triples that are evaluated as "High" by additive smoothing and Good-Turing methods are largely evaluated in the same way by humans. The triples that are evaluated as "Low" by the system however contains some triples that are evaluated as "high" by human evaluators. In general the correlation of the human evaluation and the proposed methods perform very well, with a chance of losing some good triples. This gives us a

confidence because the inferences rated as "High" by the system are usually rated as "High" by the system and a combination of the two approaches should give us a strong degree of confidence on the evaluation by the system.

We further compared the two methods to see the consistency of their results (**Fig. 3**). Both Additive and Good-Turing methods identified triples evaluated as "High" consistently with additive smoothing showing superiority in finding good inferences. However, Good-Turing smoothing shows superior quality in identifying unfamiliar inferences. One of the main concerns for Dr Inventor here is, those inferences that are rated as bad inferences may remove some creative but uncommon triples

Evaluation Results of triples using ConceptNet and WCI. Note that the global maximum association score between two concepts is 7.127 and the global minimum is 0.007. The WCI score between two words lies in the range of [0.5, 1]. Neither ConceptNet nor WCI return 0 values, so smoothing methods are not used.

Table 4. Threshold values for familiarity of COCA triples.

Score	**ConceptNet scores**	**WCI scores**
Low	$Score < 0.0345$	$score \leq 0.30$
Medium	$0.034 < score < 5$	$0.30 < score < 0.50$
High	$score \geq 5$	$score \geq 0.50$

Humans evaluated both the random triples and the familiar triples (**Table 5**). The human evaluation agrees100%) with the familiar triples as these triples are extracted from publicly available content. The human evaluation further aligns with unfamiliar triples (93%) are rated as low (according to the threshold defined in **Table 4**).

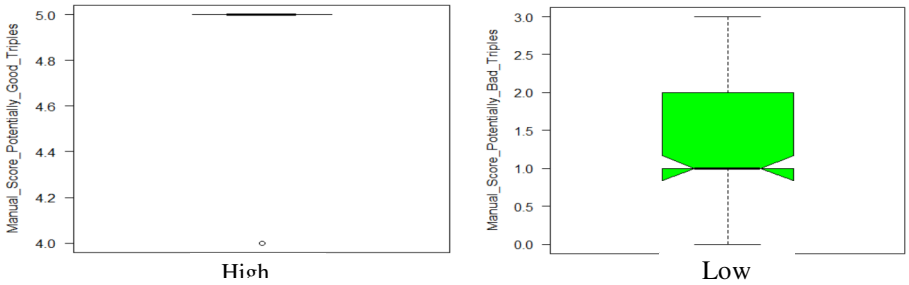

Fig. 4. Human ratings for triples considered as "High" and "Low".

It is also important to mention that (**Fig. 4**), for both methods, human scores for triples considered as "High" are overall significantly higher than human scores for triples considered as "Low". It means that both Additive Smoothing and Good-Turing are dependable at distinguishing absolutely familiar triples to humans from absolutely meaningless triples to humans. Some examples of best triples accepted by the system include "*we provide method*" and "*we show section*" and best triples rejected by the system include "*property contain penalization*" and "*millimeter be numeric*". There are also a few worst triples (e.g. "*i, k, set*") wrongly accepted by the system.

Table 5. Human evaluation of triples using ConceptNet and WCI

	Score	Triples evaluated as "high"		Triples evaluated as "Low"	
ConceptNet	High	13	46.4%	0	0%
	Medium	14	50%	1	3.4%
	Low	1	3.6%	28	96.6%
WCI	High	4	14.29	0	0%
	Medium	24	85.71%	3	10.34%
	Low	0	0%	26	89.66%

5 Conclusion

We presented our approach to evaluate the analogical inferences generated by our Dr Inventor analogical reasoning system. The subject-verb-object triples generated from the corpus were used to support an N-gram model to assess the familiarity of the novel inferences (triples) generated by the system – where familiarity was used to estimate inference validity. We further explored ConceptNet and Watson Concept Insight to evaluate these inferences. Our evaluation demonstrated that the N-gram approach is capable of differentiating good inferences from the bad ones and produced a consistent evaluation as the human evaluation. Our experimental results further shows the possibility of ranking inferences using scores generated by our methods to direct the focus of users to the most meaningful inferences. For future work, we will further explore a unified measure incorporating all three evaluation ratings to help improve the quality of inference – and the analogies that drive them.

Acknowledgement. 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

1. Simon, B., Susan, S.: Analogies in science and science teaching. Advances in Physiology Education 34(4), 167-169 (2010)
2. Gentner, D.: Structure-mapping: A theoretical framework for analogy. Cognitive Science 7, 155-170 (1983)
3. Gentner, D., Smith, L.: Analogical Reasoning. In : Encyclopedia of Human Behavior. Elsevier, Oxford, UK (2012) 130-136
4. Holyoak, K. ., Gentner, D., Kokinov, B. ., Gentner, D., Holyoak, K. .: Introduction: The place of analogy in cognition. In : The Analogical Mind: Perspectives from cognitive science. (2001) 1-19

5. Holyoak, K., Thagard, P.: Analogical mapping by constraint satisfaction. Cognitive Science 13, 295-355 (1989)
6. Keane, M., Brayshaw, M.: The Incremental Analogy Machine: A computational Model of Analogy. In : EWSL, pp.53-62 (1988)
7. O'Donoghue, D., Abgaz, Y., Hurley, D., Ronzano, F., Saggion, H.: Stimulating and Simulating Creativity with Dr Inventor. In : Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015), Park City, Utah (2015)
8. Yalemisew, A., Diarmuid, P., Donny, H., Horacio, S., Francesco, R., Dmitry, S.: Embedding a Creativity Support Tool within Computer Graphics Research. In : ECAI 2016, Workshop Modelling and Reasoning in Context (MRC), The Hague, Netherlands (2016)
9. Holyoak, K., Novick, L., Melz, E.: Component processes in analogical transfer: Mapping, pattern completion, and adaptation. In : Analogical connections. Advances in connectionist and neural computation theory 2. Ablex Publishing, Westport, CT, US (1994) 113-180
10. Alfred, J.: Language, truth and logic. Courier Corporation (2012)
11. Thomas, L.: Analogical Inference over a Common Sense Database. In : Eighteenth National Conference on Artificial Intelligence (2002)
12. Ronzano, F., Saggion, H.: Knowledge Extraction and Modeling from Scientific Publications. In : In the Proceedings of the Workshop “Semantics, Analytics, Visualisation: Enhancing Scholarly Data” co-located with the 25th International World Wide Web Conference, Montreal, Canada (2016)
13. Louise, C., Dermot, L.: Principles of Representation: Why You Can't Represent the Same Concept Twice. Topics in Cognitive Science 6(3), 390-406 (2014)
14. Veale, T., Keane, M.: The competence of sub-optimal theories of structure mapping on hard analogies. In : Proceedings of the 15th International Joint Conference on Artifical Intelligence, Nagoya, Japan, vol. 1, pp.232-237 (1997)
15. Miller, G.: WordNet: A Lexical Database for English. Communications of the ACM 38(11), 39-41 (1995)
16. Lin, D.: An Information-Theoretic Definition of Similarity., San Francisco, CA, USA (1998)
17. Cordella, L. ., Foggia, P., Sansone, C., Vento, M.: A (sub)graph isomorphism algorithm for matching large graphs. In : Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, pp.1367-1372 (2004)
18. Franceschini, M., Soares, L., Lastras, M.: Watson Concept Insights: A Conceptual Association Framework. In : Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada, pp.179-182 (2016)