Creative Puzzle Generation from Factual Content

Qing Yu, Tony Veale[1]

Abstract. Comprehensive knowledge-bases can be seen as not only rich sources of factual content – that is, answers – but also as rich sources of questions. In this paper we explore the potential of knowledge resources like the CIA World Fact book to serve as the generative basis of a series of creative educational puzzles.

1 INTRODUCTION

Large-scale knowledge-bases are, by definition, rich sources of factual content that can provide information to a variety of related applications in the same domain. This "data-store" view of knowledge-bases creates a natural inclination to perceive knowledge-bases as repositories of answers from which different applications can draw, to solve problems, to guide searches and to resolve ambiguities. But it is just as natural, albeit far less commonplace, to view knowledge-bases as sources of questions rather than facts. After all, one must know something about a domain to frame an intelligent question within that domain, and know even more to winnow and evaluate candidate answers to that question [8].

Though a meaningful perspective, this inverted view of knowledge-bases as "question-stores" may not seem an altogether useful one, until one considers the role of knowledge-bases in scholastic teaching and testing (and alternatively, in educationally-useful games that lack the medicinal taste of overtly pedagogical systems). In this context, it is more valuable to view knowledge-bases as generators of questions rather than of answers, albeit generators that are also able to answer the questions they pose. In this current work, we construe the notion of question in the broadest possible manner, to include the following kinds of query:

1. Textual questions (e.g., what, where, when, who, etc.)
2. Proportional Analogies (e.g., A is to B as C is to what?)
3. Category formation problems (e.g., odd-one-out reasoning)
4. Completion problems (e.g., what comes next, what is missing, etc.)

These queries, more properly labeled "problems" or "puzzles", run the gamut from the purely textual to the purely logical, and involve processes at every level of cognitive processing, from syntax to semantics to logical reasoning to similarity judgment to category formation. To comprehensively generate this range of problems, we require large-scale knowledge resources that can inform on a similar diversity of phenomena. These resources will range from the fully structured (e.g., relational databases) to the semi-structured (e.g., those combining flat text with an explicit structure, or text annotated with explicit mark-up tags).

In this paper we focus on just one such resource, the CIA world fact-book [4], a freely available semi-structured almanac containing a wealth of geopolitical facts. We explore the extent to which this resource can be used to generate not just questions, but puzzles that are genuinely creative [1, 2]. In the next section we consider just what this creativity entails, before we explore the puzzle potential of the CWFB in section 3, and the assessment of puzzle difficulty in section 4.

[1] School of Computer Science and Informatics, University College Dublin, Ireland. Email: {*scorain.yu, tony.veale*}@ucd.ie

2 RELATED WORK

Puzzle generation is a creative process not simply because it involves the creation of new linguistic or logical artifacts, but because such artifacts additionally imply the creation of new conceptual categories. Colton [3] notes that since many puzzles require a solver to reason about category membership, the most creative puzzles are those that hinge on the most creative categorizations of the elements in the puzzle (see also [8]). For instance, both odd-one-out puzzles and next-in-sequence puzzles require a solver to construct a category, or domain theory, to cover a set of given elements; in the latter case, this theory or category must provide a common container for all of the given elements; in the former, it must provide a compelling container for all but one, and it is this category exclusion that yields the solution. The creativity demanded of the solver must thus be matched by the generation mechanism itself.

It also follows that the interestingness of the puzzle will be a function of the interestingness of the category or theory that underpins it [6]. Again, the most interesting categories will also be the least conventional, and thus most creative, corresponding to what Barsalou [1] has dubbed "ad-hoc" categories. These are task-specific, dynamically constructed categories that draw their members from across the ranks of many different conventional categories, which makes them well-suited to creatively uniting the disparate elements of a puzzle. For example, the S.A.T.-style analogy fructose is to fruit as lactose is to what? (Answer: milk) requires the solver (and thus the generator) to construct the ad-hoc category "substances from which sugar can be extracted". The generation and solution processes for analogical puzzles [7,8] is not qualitatively different than that for odd-one-out and next-in-sequence puzzles as each necessitates the identification and construction of appropriate categories from available world knowledge.

3 GENERATING GEOGRAPHICAL PUZZLES

The CIA world fact-book (henceforth CWFB) contains sufficient geographical knowledge, expressed in a sufficiently regular format, to be viewed as a geographic knowledge-base. Viewed as such, it can support a variety of puzzle types, such as the following example of the well-known odd-one-out variety:

(1) Which of the following countries is the odd one out?
a) Belgium b) Holland c) France
d) Switzerland e) Italy

The answer here is Italy, since it is the only listed country that does not share a border with Germany. In this case, the CWFB relation borders(x, y), in combination with the CWFB entity Germany, is construed as category Countries-that-border-Germany. In this respect, puzzle generation is a process of ad-hoc category creation [1], and puzzles will be judged as creative to the extent that these categories are deemed both original and useful. While this example employs simple CWFB entities like countries, such entities can be combined to create more complex entities, like pairings of countries, as in the variation of example (2):

(2) Which of the following country pairs is the odd one out?

a) Belgium:Holland b) France:Germany
c) Belgium:Italy d) Scotland:England
e) Egypt:Sudan f) France:Belgium

Here each pairing, with the exception of 2(c), represents an instance of the relation borders(x,y). This pairwise treatment of CWFB entities can be extended to facilitate the generation of analogical puzzles, as in (3):

(3) Belgium is to France as:

a) Germany is to Holland b) Holland is to France
c) Italy is to Holland d) Germany is to Austria
e) Austria is to Hungary f) Egypt is to Sudan

The answer, 3(d), describes a pair of countries that – like the base pair – share a common border and a common language.

The CWFB contains a variety of descriptive fields for each of 231 countries – such as capital city, dominant language, religion, population, area, GDP and major industries – each of which can be used as the basis of a category-forming relation. And of course, not every puzzle need assume a multiple-choice format. The CWFB's relational contents also support guess-the-entity puzzles as in (4):

(4) Which country is a major producer of sugar and cigars, and is roughly twice the area of Delaware? [Answer: Cuba]

This puzzle type – which combines two fields from a given country description to yield a unique specification – amply demonstrates the American-centric nature of the CWFB, for in addressing its target audience of American legislators, bureaucrats and businessmen, the CWFB moulds its description of foreign countries with the cookie-cutter of American states. Though irksome, conversion to more user-centric terms is possible: if country X is A times the size of state S, and country Y is B times the size of state S, then country Y is B/S times the size of country X. By understanding the relative size of American states, a user-modelling process can simply convert any state-relative size description into a user-specific alternative (e.g., to use Spain, Belgium, etc. as units of country size).

3.1 Complex Relations, Surprise and Interestingness

The creativity of these puzzles is clearly a function of the categories that underlie them: the more unusual and interesting this category, the more challenging and creative the puzzle will be perceived to be. Each of the puzzles (1)-(4) employ an ad-hoc category in the sense that one would not expect to find such a category in a conventional taxonomization of the geographic domain. Puzzles (1) and (2) employ categories that derive from a single CWFB relation, a strategy that one can well expect to produce limited results. In contrast, puzzles (3) and (4) require a user to construct an organizing category derived from a combination of two different relations: for (3), the user must construct the category of countries that share a border and a common language, while for (4), the user must construct the intersective category of cigar producing countries twice the size of Delaware. For puzzles of type (3), this ad-hoc category may be rich in potential members, while the nature of (4) is such that this category should be a singleton.

We can expect some pairings of relations to generate more creative categories than others. While some pairings will be entirely arbitrary, the most natural (and thus, educational) combinations will pair relations that either confirm or confound a user's expectations. For instance, one can expect countries that share a border to have an increased probability of also sharing a language, and one can expect a country with a larger surface area to also have a larger population. When these expectations are violated, the combinations result in more interesting categories, such as the countries that have larger surface area but lower populations than another (E.g., Brazil and China), or smaller surface areas but considerably larger populations (e.g., China and Canada). The educational purpose of categories built upon such antagonistic combinations is to both suggest the natural inference and simultaneously demonstrate that it is not always true.

4 ASSESSING PUZZLE DIFFICULTY

A creative – as opposed to a formulaic – puzzle generator must be capable of assessing the inherent difficulty of the problems it generates. This, in effect, requires a system to possess a form of self-knowledge that is not hard-coded by its developers but which arises organically from the system itself and from the content of its underlying knowledge-sources. For instance, one abstract indicator of difficulty is the notion of familiarity: *ceteris paribus*, problems that combine unfamiliar elements should be more difficult than problems that combine familiar elements. In the case of textual, category-building puzzles, familiarity may be computationally understood in terms of statistically founded expectations such as word frequencies and age-of-acquisition statistics. In an educational context, one cannot assess puzzle difficulty independently of student knowledge, so to the extent that difficulty arises out of a lack of factual knowledge, puzzle grading is an issue of user-modeling.

Let the letters R_1, R_2, etc. denote relations. Let $f_n = R_n(x, y)$ be a fact that connects entity x to entity y via the CWFB relation R_n. For instance, one such fact is *borders(Belgium, France)*, while another is *same-language(Belgium, Holland)*. Our KB is thus a set { $R_1(...)$,$R_2(...)$,$R_3(...)$, ...}. Furthermore, let S denote the specific student to which each puzzle is addressed. We must construct a statistical model of the knowledge of S to enable the system to predict how hard the puzzles will seem to S, so that $P_S(R_n(x, y))$ is the probability that a student S will know the fact $R_n(x,y)$. Our model of S must estimate a probability for S knowing each such fact f_n in the KB, and these probabilities can be estimated in a number of ways. The first approach employs web-search, and assumes that there exists a strong correlation between the web frequencies of specific entity terms (like Belgium, Paris and Dutch) and the likelihood that a generic student will possess knowledge of these entities. Thus, common entities like France and Paris will have higher web-frequencies and higher weights, while newsworthy entities like Iraq will also have higher weights than less topical entities like Paraguay.

In the alternative student-specific approach, weights are assigned by hand by a teacher or administrator, who is responsible for initializing a user model for each student. Let $W_S(x)$ be a weight, between 0…1, assigned to a CWFB entity x (where x is a country, a city, etc.) in the model of S. These weights can be assigned to individual entities within the user-model, or to whole families of related entities simultaneously. For instance, a teacher may initialize the model for S by assigning a weight of 0.8 to Europe, 0.5 to North America, 0.4 to Africa and 0.3 to Asia; from these continent-level assignments, the system can then infer weights of 0.8 for each European country, 0.4 for each African country, and so on, unless specific overridden by specific weights to the contrary at the entity level. However weights are assigned (and we currently support both approaches), the system can estimate a base value for $P_S(R_n(x, y))$ from these weights quite simply, as follows:

$$P_S(R_n(x, y)) \quad = \quad W_S(x) \times W_S(y) \qquad (1)$$

This is a base-line probability, independent of any other knowledge S is assumed to possess (based on successes with earlier puzzles). However, geographic facts are not independent of each other, and possession of one fact should increase the likelihood of a student possessing a related fact. Thus, let $P_S(R_n(x, y) \mid R_i(x, y))$ denote the probability that S will possess the fact $R_n(x, y)$ if S already possesses the fact $R_i(x, y)$. Let $P_S(R_n(x, y) \mid R_i(x, y), R_j(x, y))$ denote the probability of S knowing $R_n(x, y)$ if S already knows both $R_i(x, y)$ and $R_j(x, y)$. In the absence of other information, the probability that a student will infer that two adjacent countries will share the same language is dependent on the frequency with which adjacent countries are observed to share the same language in the CWFB. These probabilities can thus be estimated by calculating the co-occurrence of relations in the CWFB as follows:

$$P_S(R_n(x, y) \mid R_i(x, y)) \quad = \quad |\{R_n\} \cap \{R_i\}| \,/\, |\{R_i\}| \qquad (2)$$

where $\{R_i\}$ denotes the extension of the relation R_i, that is, the set of entity pairings in the CWFB to which the relation R_i can be applied to produce a valid geographic fact. Likewise,

$$P_S(R_n(x, y) \mid R_i(x, y), R_j(x, y))$$
$$= \quad |\{R_n\} \cap \{R_i\} \cap \{R_j\}| \,/\, |\{R_i\} \cap \{R_j\}| \qquad (3)$$

Of course, the difficulty of a multiple-choice puzzle is not only a function of the likelihood that a student S will know the right answer, but also a function of the likelihood that S will believe a competing distractor to be correct when it is in fact false. The more plausible a distractor, given the knowledge possessed by S, then the harder the choice faced by S and thus, the harder the puzzle is perceived to be. We can calculate $P_S(R_n(x, y))$ when $R_n(x, y)$ is a false distractor as follows:

$$P_S(R_n(x, y)) \quad = \quad \max_{i,j}(P_S(R_n(x, y) \mid R_i(x, y), \; R_j(x, y)) \times W_S(x) \times W_S(y)) \qquad (4)$$
$$= \quad \max_{i,j}((|R_n \;\&\; R_i \;\&\; R_j| \times W_S(x) \times W_S(y)) \,/\, |R_i \;\&\; R_j|) \qquad (5)$$

that is, by finding the pair of facts $R_i(x, y)$ and $R_j(x, y)$ known by S that are most likely to make S conclude that the erroneous assertion $R_n(x, y)$ is in fact true. Thus, if $R_n(x, y)$ denotes the answer of a puzzle Z, and $R_i(a, b)$ denotes the most believable distractor (to S) of Z, whether a true or false state of affairs, then

$$\text{difficulty}_S(Z) \;\; \propto \;\; P_S(R_n(x, y)) \times P_S(R_i(a, b))^{-1} \qquad (6)$$

5 PRELIMINARY EVALUATION

In describing 231 different countries, the CWFB affords us 9 primitive geographical relationships for constructing puzzle-specific categories: shares-border, same-religion, same-ethnicity, same-language, same-currency, same-continent, bigger/small/same area and population, and same-landlocked-status. The converse of these relations (no-shared-border, different-religion, etc.) also afford us 9 kinds of primitive distractor relationship. However, since these primitive relationships provide the most conventional means of comparing countries, they are also the least challenging, and thus least interesting, from a puzzle perspective. We can combine these primitive relationships to create more interesting composites, as illustrated in Table 1.

Accordingly, there are 118^2 analogies of the form A:B::C:D employing relation 1 (both pairs share a border and speak the same language), and just 36^2 using relationship 3. Composite relationships hold multiple advantages over their primitive ingredients: for one, they yield less obvious, more ad-hoc categories that in turn yield more interesting and atypical puzzles; for another, a composite relationship $R_1 \wedge R_2$ gives rise to more subtle distractors, drawn from the categories $R_1 \wedge \neg R_2$ and $\neg R_1 \wedge \neg R_2$. These distractors, and their instantiations in the CWFB, are enumerated in Table 2.

Of course, some combinations of primitive relationships are more atypical, and thus more interesting, than others. For instance, one expects countries with bigger landmasses to have larger populations, but this expectation is not always realized, as exemplified by the pairing of China and Canada. In these cases, the solution category is constructed around the relationship $R_1 \wedge \neg R_2$ when the relationship $R_1 \wedge R_2$ is considered more likely; more formally, $P_S(R_2(x, y) \mid R_1(x, y)) > P_S(\neg R_2(x, y) \mid R_1(x, y))$.

Table 1. Composite relationships in CWFB

	Relation Combination	Number of matching country pairs
1	A and B share a border and have the same dominant religion	118
2	A and B share a border and have the same dominant ethnicity	24
3	A and B share a border and have the same dominant language	36
4	A and B share a border in different continents	16
5	A and B share a border and have the same landlocked status	43
6	A and B share a border and have the same currency	40
7	A and B share the same dominant religion and ethnicity	298
8	A and B share the same dominant religion and language	244
9	A and B share a dominant ethnicity in different continents	245
10	A and B share the same dominant ethnicity and language	46
11	The area of A is greater than B but A has a smaller population	348

Table 2. Composite Distractors for Composite Relationships in CWFB

Composite Distractor	Number
{neighbor, different religion}	129
{neighbor, different ethnicity}	105
{neighbor, different language}	49
{neighbor, different currency}	62
{neighbor, differently landlocked}	257
{neighbor, same continent}	297
{same religion, different ethnicity}	1148
{same religion, different language}	1070
{same ethnicity, same continent}	222
{same ethnicity, different language}	8
{larger area, bigger population}	1819
{larger area, smaller population}	348
{not neighbor, same religion}	2660
{not neighbor, same ethnic}	241
{not neighbor, same language}	837
{not neighbor, same currency}	357
{not neighbor, same landlocked}	4204
{different ethnicity, same continent}	5817
{different religion, same ethnicity}	55
{different religion, same language}	128
{different ethnicity, same continent}	1583
{not neighbor, same ethnicity}	287
{smaller area, smaller population}	1819
{smaller area, bigger population}	348

6 CONCLUSION

Colton [5] notes that scientific advances sometimes occur by specifying exactly the right question to ask at the right time, before then finding an answer. In this view, problematic questions can be every bit as creative as the solutions they elicit. To this end, this current work represents an inversion of the conventional logic regarding the exploitation of large-scale knowledge-resources, in which every fact in a KB is an over-specified question that contains its own answer. As such, if suitably generalized, each fact becomes a question waiting to be answered. The CWFB is just one factual resource that can be exploited for such ends. Though much work remains to be done even with this relatively small-scale knowledge compendium, issues of interestingness, atypicality, creativity and difficulty can all be richly explored within its confines.

ACKNOWLEDGEMENTS

We would like to thank Enterprise Ireland for supporting this research through a grant from the Commercialization Fund.

REFERENCES

[1] Barsalou, L. (1983). Ad-Hoc Categories. Memory and Cognition, 11(3), 211-227.

[2] Boden, M. (1990). The Creative Mind: Myths and Mechanisms. London: Weidenfeld and Nicholson.

[3] Boden, M. (1999). Computational models of creativity. Handbook of Creativity, pp 351–373.

[4] The CIA World Fact-book (2006). Available from: http://www.cia.gov

[5] Colton, S. (2002). Automated Puzzle Generation. In Proceedings of the AISB'02 Symposium on AI and Creativity in the Arts and Science, 2002.

[6] Colton, S., Bundy, A. and Walsh, T. (2000). On the Notion of Interestingness in Automated Mathematical Discovery. International Journal of Human Computer Studies, Volume 53. No. 3, 351-375.

[7] Turney, P and Littman. M. (2003). Learning Analogies and Semantic Relations. Technical Report ERB-1103, National Research Council, Institute for Information Technology.

[8] Veale, T. WordNet sits the S.A.T.: A Knowledge-Based Approach to Lexical Analogy. In the proceedings of ECAI'2004, the 16th European Conference on Artificial Intelligence. John Wiley: London.