On Creative Uses of Word Associations

Hannu Toivonen[1], Oskar Gross[1], Jukka M. Toivanen[1], and Alessandro
Valitutti[1]

Abstract The ability to associate concepts is an important factor of creativity. We investigate the power of simple word co-occurrence analysis in tasks requiring verbal creativity. We first consider the Remote Associates Test, a psychometric measure of creativity. It turns out to be very easy for computers with access to statistics from a large corpus. Next, we address generation of poetry, an act with much more complex creative aspects. We outline methods that can produce surprisingly good poems based on existing linguistic corpora but otherwise minimal amounts of knowledge about language or poetry. The success of these simple methods suggests that corpus-based approaches can be powerful tools for computational support of creativity.

1 Introduction

The ability to associate concepts, ideas, and problems is an important factor of creativity. Creative people often are able to see or establish connections and analogies where others could not, and this ability may lead to better solutions to problems or new pieces of art.

We are interested in using computers to support or even accomplish tasks involving verbal creativity. In this paper, we will more specifically look at methods that use word associations derived from word co-occurrences in large corpora. For instance, words 'hand' and 'fist' occur relatively often together, indicating that they are semantically related.

More specifically, our goal is to explore the power of word co-occurrences on tasks that require lexical creativity. We keep all other linguistic and world knowledge at a minimum to test how far plain word associations can take

Department of Computer Science and HIIT, University of Helsinki, Finland
`firstname.lastname@cs.helsinki.fi`

us. On the other hand, the methods are less dependent on any particular language and resources.

We address two specific tasks. The first one is the taking the Remote Associates Test [10], a psychometric test of creativity. It directly measures the ability to associate words. The second task is generation of poetry, an act with much more complex creative aspects. Also in this case, word associations can be used as a key component of a poetry generation system.

This paper is structured as follows. We first review some background in Section 2. We then address the Remote Associates Test of creativity in Section 3 and generation of poems in Section 4. We conclude in Section 5.

2 Background

We next provide a brief background for word associations: first the RAT creativity test and then word co-occurrence measures.

Remote Associates Test The Remote Associates Test (RAT) measures the test subject's ability to find associations between words. In the test, three unrelated *cue words* are presented to the subject, e.g., 'thread', 'pine', and 'pain'. The person then tries to identify a fourth word, the *answer word*, which is related to each of the cue words. In this example, the solution is 'needle'.

The Remote Associates Test was developed by Mednick [10] in the 1960s to test creativity defined as "*the forming of associative elements into new combinations, which either meet specified requirements or are in some way useful*". The test is frequently used by psychologists even if some argue that it is not a good measure of creativity.

In practice, RAT measures the ability to discover new associations between concepts that are not typically connected. Performance on RAT also relates to how well one can generate original ideas [5].

Log-likelihood Ratio We now describe how we use log-likelihood ratio (LLR) to measure how strongly two words are related in a give corpus. We assume a corpus of unstructured documents, and we treat documents as bags of sentences and sentences as bags of words. Instead of sentences, we can consider all n-grams, i.e., sequences of n consecutive words.

The LLR as applied here is based on a multinomial model of co-occurrences of words (see, e.g., Dunning [4]). The multinomial model of any pair $\{x, y\}$ of words has four parameters $p_{11}, p_{12}, p_{21}, p_{22}$, corresponding to the probabilities of events $\{x, y\}$, $\{\neg x, y\}$ $\{x, \neg y\}$ $\{\neg x, \neg y\}$. The ratio of likelihoods of two multinomial models is computed, a null model and an alternative model. The null model assumes independence of words x and y. Their probabilities are estimated as their frequencies in the data, and the probabilities of their different combinations $(p_{11}, \ldots, p_{22})$ are obtained by simple multiplication

(assuming independence). The alternative model, in turn, is the maximum likelihood model which assigns all four parameters from their observed frequencies.

The log-likelihood ratio test is then defined as

$$LLR(x, y) = -2 \sum_{i=1}^{2} \sum_{j=1}^{2} k_{ij} \log(p_{ij}^{null}/p_{ij}), \tag{1}$$

where k_{ij} are the respective counts. LLR measures how much the observed joint distribution of words x and y differs from their distribution under the null hypothesis of independence, i.e., how strong the association between them is in the given corpus.

Related work Literature on measuring co-occurrences or collocations of words is abundant. Standard techniques include the following.

Log-likelihood ratio is a non-parametric statistical test often used for co-occurrence analysis [4]. Unlike some other measures, log-likelihood ratio does not overestimate the importance of very frequent words.

Latent Semantic Analysis [3] aims to find a set of concepts (instead of terms) in a corpus using singular value decomposition. The semantic similarity (relatedness) of two words can then be estimated by comparing them in the concept space. Latent semantic analysis has then evolved to *Probabilistic Latent Semantic Analysis* [8] and later to *Latent Dirichlet Allocation* [1].

We are also interested in building *networks of word associations*. Concepts maps, mind maps, and mental maps are some well-known examples of specific types of networks designed to help learning and creativity or to model subjective information processing. As an example of work in this area, Tseng et al. [15] proposed a two-phase concept map construction algorithm which uses fuzzy sets and multiple types of rules to generate concept maps.

3 Solving the Remote Associates Test of Creativity

We now illustrate the power of simple word co-occurrence analysis for the RAT test of creativity [7]. This is, admittedly, a narrow and specific context. However, if the human capability to perform well in RAT is related to creativity, then certainly the capability of a computer performing well is an encouraging indication of its ability to potentially perform creative tasks, or at least to help humans in tasks requiring creativity. The more complex task of creating poetry will be addressed in the next section.

Data We used 212 RAT items of Bowers *et al.* [2] and Mednick & Mednick [11], divided to a training set of 140 items and a test set of 72 items. As a corpus, we use Google 2-grams [12]. We removed stopwords, i.e., common and therefore uninformative English words, using the NLTK stopword list.

3.1 Modeling RAT items computationally

Let quadruple $r = (c_1, c_2, c_3, a)$ denote a RAT item, where c_i is the ith cue word and a is the answer word. In a probabilistic formulation, the task is to predict the most likely answer word a given cue words c_1, c_2, c_3. Assuming independence between the cue words, i.e., using the Naïve Bayes model, we obtain

$$P(a|c_1, c_2, c_3) \propto P(a, c_1, c_2, c_3) = P(a) \prod_{i=1}^{3} P(c_i|a). \qquad (2)$$

We estimate the (conditional) probabilities from the relative frequencies of the words in the Google 2-grams, and find the word a that maximizes Eq. 2. For more details, see Gross *et al.* [7].

The problem is challenging. There are millions of words to choose from, and even when only considering words that co-occur with each of the cue words, there are thousands of possibilities.

3.2 Experiments

When tested on the RAT items from psychometric literature, the above model provided the correct answer in 66% of cases both in the training and the test sets. Clearly, computers can perform well in such limited tests of creativity by simple co-occurrence analysis even if the search space is very large.

Looking at the 33% of unsuccessful cases, the system often answered with a plural form when the correct answer was singular. Additionally, in some of the test items, a cue word does not occur in the 2-grams at all as an individual word, but only as part of a compound word (with the answer word, for instance). Obviously, one could engineer the method to deal with such issues with plurals and compound words, but the main point is already clear: the performance of the system is better than that of an average person. Item-wise solution rates are typically 30–70%, so the performance of 66% correct solutions can actually be considered very good. This is a clear indication that computers can solve some tasks that are considered to require creativity.

4 Creation of poetry

We now move on to a much more demanding creative task, writing of poems. We outline a corpus-based approach for this task; more details are given by Toivonen *et al.* [14].

In the literature, several different methods and systems have been proposed for poetry generation (e.g., [9, 6, 16, 13]). They use, among others, statistical

approaches, case-based reasoning, and evolutionary algorithms. Many of the best performing systems are based on explicitly coded knowledge about the world (e.g., using formal logic) as well as rich linguistic knowledge (e.g., a generative grammar or a tagged corpus of poetical text fragments). A different family of approaches is based on Markov chains or n-grams. They learn a model of word sequences from a given corpus and use this model to produce new poetry. The typical shortcoming of such approaches is that longer sequences of text make no sense grammatically or semantically.

Our goal is to minimize all explicit knowledge about the world or the language, and instead rely on given corpora for implicit knowledge about them. Additionally, some off-the-shelf linguistic analysis tools are needed (lemmatizer, part-of-speech tagger, morphological analyzer and synthesizer). We take corpora as input, just like Markov models, but the method is completely different.

We use two corpora. The first one, called *background corpus*, is used to analyze word co-occurrences and to construct a word association network. This network is used to control the topic and semantic coherence of poetry. The second corpus, called *grammar corpus*, is used as a set of grammatical examples or templates in an instance-based manner.

Data We currently generate poetry in Finnish. The background corpus is Finnish Wikipedia, and the grammar corpus consists of older Finnish poetry.

4.1 Method

The input to the method essentially consists of three items: the background corpus, the grammar corpus, and a topic word.

The contents and coherence of the poem are controlled by using words that are related to the given topic word in the background corpus, as measured by LLR. The grammatical correctness, in turn, is partially guaranteed by taking a random fragment (e.g., a sentence or a poem) from the grammar corpus, and using its grammatical structure in the generated poem.

More specifically, an example fragment of the desired length is chosen from the grammar corpus. It is then analyzed morphologically for the part of speech, case, verb tense, clitics, etc. of each word.

Then, words (especially verbs, nouns, adjectives and adverbs) in the fragment are substituted independently, one by one, by words associated with the given topic. The substitutes are of the same type with the original words and are transformed to similar morphological forms. The original word is left intact, however, if there are no words associated with the topic that can be transformed to the correct morphological form. This can happen, e.g., if the morphological form is rare or complex.

4.2 Results

We next give some example poems generated by the method, translated from Finnish originals. The first poem is about children's play (in the left column). The original text on which it is based (in the right column) is a fragment of a poem by Uuno Kailas.

Computer-generated poem	Text used as a template
How she played then	how she played once
in a daring, daring whispering	in a big green park
under the pale trees.	under the lovely trees.
She had heard for fun	She had watched for fun
how her whispering	how her smile
drifted as jingle to the wind.	fell down as flowers,

The following poem is about hand. The poem fragment used as a template is by Edith Södergran.

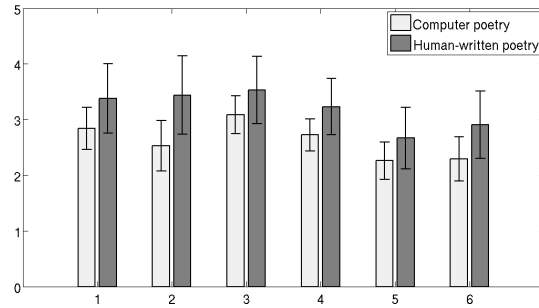| Computer-generated poem	Text used as a template
In a pale fist	In a gloomy forest
in a well-balanced fist,	In a dim forest
the buds are so pale	flowers are so pale
in your image lies a dear child god.	In the shadow lies a sick god

The last example is about snow. The text used as a template is by Eino Leino.

Computer-generated poem	Text used as a template
Lives got the frolic ways,	Waves fared the wind's ways,
snow the home of time,	sun the track of time,
softly chimed abandoned homes,	slowly skied for long days,
softly got frolics beloved –	slowly crept for long nights –
ripening crop got the snows' joys.	day wove the deeds of moons

We evaluated the poetry using a panel of twenty random subjects. Each of them evaluated 22 poems, of which 11 were computer-generated and 11 human-written. The poems were presented in a random order and the subjects were not informed that some of the poems are computer-generated. Each poem was evaluated qualitatively along six dimensions: (1) How typical is the text as a poem? (2) How understandable is it? (3) How good is the language? (4) Does the text evoke mental images? (5) Does the text evoke emotions? (6) How much does the subject like the text? These dimensions were evaluated on the scale from one (very poor) to five (very good).

On each of the dimensions, the 67% confidence intervals of the answers for computer-generated vs. human-written poetry overlap a lot (Figure 1), indicating that a large fraction of computer-generated poetry is as good as human-written poetry, even if on average human-written poetry is better.

Fig. 1 Subjective evaluation of computer-generated and human-written poetry along six dimensions (see text). Results are averaged over all subjects and poems; whiskers include one standard deviation above and below the mean.

This is a striking result given the simplicity of the methods, and again indicates that simple text analysis methods can be powerful components of verbally creative systems.

5 Conclusions

We have shown how word co-occurrence analysis can be used to perform acts requiring verbal creativity. The Remote Associates Test directly measures the capability to associate words, which is a relatively easy task for a computer when it is given a large corpus. Generation of poetry is a much more complex problem, but word associations together with existing poetry as templates can give surprisingly good results.

The results indicate that word co-occurrence analysis can be a powerful building block of creative systems or systems that support human creativity. While 2-grams were sufficient for achieving a high score on RAT, more relaxed co-occurrences are likely to provide more interesting semantic associations to support or inspire creativity, as suggested by Gross *et al.* [7].

We have used statistical, co-occurrence-based associations of words. The benefit is that their coverage is large, but at the same time they lack explicit semantics. Our results on computational generation of poetry [14] show that this does not prevent them from being used in tasks that demand higher verbal creativity.

In this paper, we have only touched on some specific problems in verbal creativity. We believe that corpus-based approaches can be powerful for many other creative problems, too: they are adaptive and the methods are largely independent of language and resources such as lexicons or knowledge-bases.

References

1. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. Journal of Machine Learning Research 3:993–1022
2. Bowers K, Regehr G, Balthazard C, Parker K (1990) Intuition in the context of discovery. Cognitive psychology 22(1):72–110
3. Deerwester S, Dumais S, Furnas G, Landauer T, Harshman R (1990) Indexing by latent semantic analysis. Journal of the American society for information science 41(6):391–407
4. Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. Computational linguistics 19(1):61–74
5. Forbach G, Evans R (1981) The remote associates test as a predictor of productivity in brainstorming groups. Applied Psychological Measurement 5(3):333–339
6. Gervás P (2001) An expert system for the composition of formal spanish poetry. Journal of Knowledge-Based Systems 14(3–4):181–188
7. Gross O, Toivonen H, Toivanen JM, Valitutti A (2012) Lexical Creativity from Word Associations. In: Proceedings of Knowledge, Information and Computational Creativity Support Systems, Melbourne, Australia, accepted for publication
8. Hofmann T (1999) Probabilistic latent semantic indexing. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY, USA, pp 50–57
9. Manurung HM, Ritchie G, Thompson H (2000) Towards a computational model of poetry generation. In: Proceedings of AISB Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science, pp 79–86
10. Mednick S (1962) The associative basis of the creative process. Psychological review 69(3):220
11. Mednick S, Mednick M (1967) Examiner's Manual, Remote Associates Test: College and Adult Forms 1 and 2. Houghton Mifflin
12. Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, Team TGB, Pickett JP, Hoiberg D, Clancy D, Norvig P, Orwant J, Pinker S, Nowak MA, Aiden EL (2011) Quantitative analysis of culture using millions of digitized books. Science 331(6014):176–182
13. Netzer Y, Gabay D, Goldberg Y, Elhadad M (2009) Gaiku : Generating haiku with word associations norms. In: Proceedings of NAACL Workshop on Computational Approaches to Linguistic Creativity, pp 32–39
14. Toivanen JM, Toivonen H, Valitutti A, Gross O (2012) Corpus-based generation of content and form in poetry. In: International Conference on Computational Creativity, Dublin, Ireland, pp 175–179
15. Tseng SS, Sue PC, Su JM, Weng JF, Tsai WN (2007) A new approach for constructing the concept map. Computers and Education 49(3):691–707
16. Wong MT, Chun AHW (2008) Automatic haiku generation using VSM. In: Proceedings of ACACOS, pp 318–323