The Word-Weaving Clock: Time Constraints in Word Associations

Alessandro Valitutti
Phedes Lab
http://phedes.com
alessandro.valitutti@phedes.com

Abstract

A wise design of time constraints is crucial for the computational realization of several creative tasks. In this paper we focus on live generation of word associations unfolding semantic paths modulated by contextual clues. We present the *Word-Weaving Clock*, a lexical creative system performing the generation of word associations where both semantic and time constraints are taken in account. The system is meant to be used interactively as part of a live demonstration.

Introduction

Do time constraints promote creativity? And if it is true for human creativity, does this also apply to the computational one?

A good number of computational-creativity tasks, mostly inspired by Margaret Boden's seminal ideas (Boden, 1990), are defined as a search in a conceptual space. Unfortunately, search processes are time-consuming, especially in those cases that require access to a vast amount of common sense knowledge. This affects the feasibility of tasks in which a time-constrained performance is essential for the appreciation of creativity. Let's think, for example, of a musical jam session, freestyle rap, or poetry slam. There are forms of creative brainstorming and stand-up comedy where prompt responsiveness or interactivity are crucial aspects of the game.

In this paper, we focus on a specific task consisting of a live generation of word associations unfolding semantic paths modulated by contextual clues. We present the *Word-Weaving Clock*, a lexical creative system performing the generation of word associations where both semantic and time constraints are taken into account. Moreover, we report the conceptual elements and the resources employed in the design and implementation of the system. The task and the resources upon which it was built have been kept as simple as possible, to facilitate its performance replicability and its potential use as a component of more complex systems. Finally, we give examples of outputs suggesting how the system would behave as part of a live presentation.

Background

As reported by Haught and Johnson-Laird (2003), "constraints are at the heart of the creative process. They govern the generation of ideas". In particular, time constraints affect creative writing (Biskjaer et al., 2019). They should be tuned carefully because "creativity can be compromised by both scarcity and abundance of time" (Liikkanen et al., 2009; Baer and Oldham, 2006).

A major issue in applying time constraints is that search processes are necessarily time-consuming. There is no way to reduce arbitrarily the time required to perform knowledge discovery in a large dataset. For this reason, indexing is an essential aspect of the design of an information retrieval algorithm, as in the case of search engines (Zuze and Weideman, 2011). In the recent development of transfer learning, knowledge and language models can be effectively reused in a way to reduce machine-learning training time (Zhuang et al., 2021). Since both writing and ideation processes are based on the discovery and exploitation of links between concepts, computational creativity efforts has been put into lexical associations. Gross et al. (2012) focused the Remote Association Test, a task in which, given three words, the word semantically connecting all of them is required. They implemented it through co-occurrence frequencies of word pairs in a large textual corpus.

In our version of this task, we have two input words with different semantic roles. The *main word* generates a set of candidate words according to the relation modeling associations in the common-sense knowledge. The *clue word* – introduced either at the beginning or, interactively, at any point in the process – provides a semantic context and allows the system to make a selection from the candidate words. For instance, the main word *'eyes'* generates *'skin'* according to the *'body'* clue word, and *'night-sight'* according to the *'pleasure'* clue word. Most importantly, a *one-second time constraint* is imposed so that a new output word is generated every second. The overall aimed effect is providing creativity both in the interactive experience and in the generated path of word associations. This periodic recursion in the generation of word associations is meant to evoke the flow of consciousness emerging as blending of semantic pulses, as an inspiration from the notion of Damasio's core consciousness, which is seen as "created in pulses, each pulse triggered by each object that we interact with or that we recall" (Damasio, 1999).

Task Description

The task consists of the following elements:

- **Word Selection Step.** Firstly, the exploration of semantic relatedness, modeling associations in common-sense knowledge, allows the system to identify a set of candidate words related to the input word. Next, a further contextual semantic constraint (the *semantic slanting*) allows the system to select the *output word*.
- **One-Second Time Constraint.** It is prefixed as the one-second time interval within which the word selection step should be performed.
- **Iteration.** The word selection step is iterated so that the output word became the input word of the next step, thus generating an associative word path as the overall product of the interaction. To avoid repetitions, the words in the associative paths are removed from the candidate words in the next word selection step.

As an example of output, the word *'travel'* is associated, in the system, to the word path *"booking → hotel → casino-hotel → jack-in-the-box → ..."*. With the clue *'happy'*, the path is *"go around → make up → know how → well-wishing → ..."*, while with the clue *'unhappy'*, the path is *"go around → go slow → long-suffering → ill-being → ..."*. Finally, the input word could be used as the clue word of itself, so reinforcing association closer to its semantic domain, e.g. producing *"booking → tour → visiting → shopping → ..."*.

Implementation

1. Common-Sense Associations: Word Embeddings

One of the most effective ways to implement semantic relatedness of words (the so-called "word similarity") consists of the cosine distance between word pairs represented as vectors (Mikolov et al., 2013). To measure word similarities we employ word embedding provided by *Spacy*[1], an open-source software library in Python for advanced natural language processing (Hiippala, 2021; Jurafsky and Martin, 2000). In particular, we use word2vec model for word embedding Jatnika, Bijaksana, and Suryani (2019) trained it on a large-scale language model in English[2]. In the analysis of word similarity, the procedure filtered word pairs with similarity values greater or equal to 0.2.

2. Semantic Slanting: Clue Word

Once found a set of candidate words, all semantically related to the current word, the further selection is performed according to the semantic relatedness with the clue word slanting the search of associations toward a specific semantic domain or connotation. The clue word could represent an emotion or, more abstractly, a sentiment polarity (e.g., either positive or negative). In a possible live demonstration, the system is meant to insert or modify the clue word at runtime, in such a way as to modulate the associative path interactively.

[1]https://spacy.io
[2]spacy.io/models/en#en_core_web_lg

3. Time Constraint: Indexing

We need to reduce the running time for each word selection step below a single second. The main bottleneck is measuring word similarity for all word pairs whose first term is the input word. Word similarity search is highly time-consuming. For instance, measuring word similarity with Spacy takes an average time of 20 milliseconds (with a CPU clock speed of 3.2 GHz). So measuring similarity with 3000 words is sufficient to exceed the one-second time constraint. Oxford Dictionary has 273,000 headwords, 71,476 of them being in current use[3]. So if we want to compare an input word with all headwords, the running time would be more than one hour and a half. Therefore, it is clear that only smartly-designed indexing allows the system to satisfy the one-second time constraint.

The following points summarize the choices that allowed us to satisfy the time constraint:

1. We collected from the Web about 27600 English nouns. We used WordNet as a tool for selecting nouns (Miller, 1995).
2. Next, we randomized the list of nouns and partitioned it into a number of sublists of the size of about 3000 items, then we calculated mutual similarities inside each sublist. In this way, we were able to develop the full system (with the time constraint) with an increasing efficiency according to the current state of indexing.
3. Finally, we calculated similarities between different sets, adding a subset a time until all mutual similarities were indexed.

In general, the number of similarity measurements N_{sm} for a word set of size n is:

$$N_{sm} = \frac{(n-1)\cdot n}{2} - 1$$

Therefore, the previewed overall running time for the 12000 nouns plus three subsets of 3000 items takes about three weeks of computation. The full set of 27000 nouns would need 84 days of computation.

As part of this research contribution, we provided the indexed resource in a text file[4]. To implement fast indexing and retrieval, we created a database using sqlite3[5] – Python interface to SQLite[6] database engine library. In the current version of the *Word-Weaving Clock*, most of the contribution to running time is mostly consisting in searching the table rows with similarity values, having the input word as either the first or the second element of the word pair.

[3]https://en.wikipedia.org/wiki/List_of_dictionaries_by_number_of_words – retrieved May 28, 2022.
[4]https://www.kaggle.com/datasets/alessandrovalitutti/noun-similarity-pairs
[5]https://docs.python.org/3/library/sqlite3.html
[6]https://www.sqlite.org/index.html

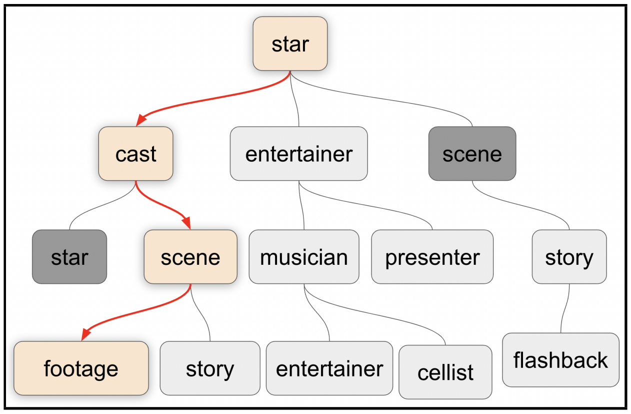

Figure 1: Example of similarity word tree as the composition of indexed word-similarity associations.

Examples of Outputs

Figure 1 shows an example of indexed common-sense word associations (without semantic slanting). Starting from the root word *'star'*, the procedure selects the node containing the word most similar to the word in the current parent node. Dark grey nodes contain words already included in the associative path, thus removed from the future selections.

Table 1 compares different possible associative paths generated from the word *'star'* and corresponding to different clue words (from the second to the last column) or without semantic slanting (first column). Clue words with opposite polarity (e.g., *'success'* vs. *'failures'*) modulate the generation of paths with positive and negative sentiment, respectively (columns two and three). In particular, emotion words with opposite polarity (e.g., *'joy'* vs. *'disgust'*) allow the procedure to generate paths characterized by affective valence (columns four and five). Moreover, using a domain word as a clue word (such as *'physics'* or *'movie'* induces a reinforcement in the generation of words in that domain (see columns six and seven). Finally, comparing columns one (no clue word) and seven (*'movie'* as clue word), we can see that semantic slanting tends to keep word associations in the same domain, although without semantic slanting the path can include more domains with higher semantic cohesion (e.g. *cinema* and *music* in column 1).

We emphasize that the reported examples are generated with a preliminary version of the similarity database. Next versions will access associations with higher values of word similarity and corresponding quality in the semantic cohesion.

Conclusions

The *Word-Weaving Clock* has been conceived primarily to stimulate interest in the study of timing in computational-creativity tasks. Time constraints are meant to be taken into account not only to improve user experience but also the design process itself since it challenges the designer to perform a wise balance between offline indexing and runtime running. Although the main intended creative value is in the interactivity experience that comes from a live demonstration of the system, the produced associative path can be considered as an artifact exhibiting creative value per se. A testbed version of the system could be used for the offline exploration of alternate paths according to different semantic and temporal parameters, for identifying values and ranges useful to improve the interactive version. The provided dataset of similarity values on English nouns is a potential handful resource for allowing researchers to further explore common-sense associations without the need for indexing.

As a possible application, time beat synchronized associative paths could be used as a backbone for the real-time selection of tweets or poetic lines. Semantic slanting could be performed according to more complex semantic patterns modeling personality traits. For example, the alternation of positive and negative slanting words could be used to mimic emotional instability.

Our next step in the development of the proposed system is to explore processing threads, to make it capable of performing both information indexing and retrieval concurrently. In some contexts, it would be interesting providing the system with full autonomy in analyzing a set of texts and building its model of word embedding, according to which the word associations will be discovered. Finally, we intend to explore different scenarios and user interfaces to make the system useful as a testbed for the offline exploration of associative paths.

Author Contributions

A.V. ideated and wrote the paper alone.

Acknowledgements

This research has received no external funding.

References

Baer, M., and Oldham, G. R. 2006. The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology* 91(4):963–970.

Biskjaer, M. M.; Frich, J.; Vermeulen, L. M.; Remy, C.; and Dalsgaard, P. 2019. How time constraints in a creativity support tool affect the creative writing experience. In *Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE 2019)*, 100–107.

Boden, M. A. 1990. *The Creative Mind*. London: Abacus.

Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

Gross, O.; Toivonen, H.; Toivanen, J. M.; and Valitutti, A. 2012. Lexical creativity from word associations. In *International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 35–42.

Haught, C., and Johnson-Laird, P. N. 2003. Creativity and constraints: The production of novel sentences. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 25, 528–532.

	success	**failure**	**joy**	**disgust**	**physics**	**movie**
cast	role	role	glamour	scene	role	flick
scene	success	concern	thrill	horror	general	soundtrack
footage	progress	failure	enjoyment	trepidation	assumption	trailer
trailer	confidence	fault	happiness	distaste	terms	scene
soundtrack	dedication	integrity	elation	exasperation	language	horror
intro	attention	disregard	emotion	contempt	arithmetic	story
interlude	influence	unwillingness	tears	stupidity	quantum	footage
accompaniment	concern	inefficiency	solace	rudeness	geometry	watching
choral	failure	reluctance	tranquility	unwillingness	singularity	streaming
oratorio	integrity	inconsistency	joyfulness	uneasiness	neutrino	sex
castrato	sustainability	assumption	faithfulness	elation	climatology	blonde
mezzo	partnership	burden	dedication	emotion	oceanography	girlfriend
pianissimo	milestone	severity	confidence	fascination	anthropology	thought
cadenza	challenge	condition	vitality	anticipation	lecturer	sort
hapsichord	make	necessity	imagination	anxiety	grad	make

Table 1: Example of associative word paths generated from from the input word *'star'* and modulated by different clue words.

Hiippala, T. 2021. Applied Language Technology: NLP for the Humanities. In *Proceedings of the Fifth Workshop on Teaching NLP*.

Jatnika, D.; Bijaksana, M. A.; and Suryani, A. A. 2019. Word2Vec Model Analysis for Semantic Similarities in English Words. *Procedia Computer Science* 157:160–167.

Jurafsky, D., and Martin, H. J. 2000. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, N.J.: Prentice Hall.

Liikkanen, L. A.; Björklund, T. A.; Hämäläinen, M. M.; and Koskinen, M. P. 2009. Time constraints in design idea generation. In *Proceedings of the International Conference on Engineering Design (ICED'09)*.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13)*, volume 2, 3111–3119.

Miller, G. 1995. Wordnet: a lexical database for english. *Communications of the ACM* 38(11):39–41.

Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; and He, Q. 2021. A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109(1):43–76.

Zuze, H., and Weideman, M. 2011. A comparative analysis of search engine indexing time. In *Proceedings of the 13th Annual Conference on World Wide Web Applications*.