

MERALI at SemEval-2017 Task 2 Subtask 1: a Cognitively Inspired approach

Enrico Mensa
University of Turin, Italy
Dipartimento di Informatica
mensa@di.unito.it

Daniele P. Radicioni
University of Turin, Italy
Dipartimento di Informatica
radicion@di.unito.it

Antonio Lieto
University of Turin, Italy
ICAR-CNR, Palermo, Italy
lieto@di.unito.it

Abstract

In this paper we report on the participation of the MERALI system to the SemEval Task 2 Subtask 1. The MERALI system approaches conceptual similarity through a simple, cognitively inspired, heuristics; it builds on a linguistic resource, the TTCS^ε, that relies on BabelNet, NASARI and ConceptNet. The linguistic resource in fact contains a novel mixture of common-sense and encyclopedic knowledge. The obtained results point out that there is ample room for improvement, so that they are used to elaborate on present limitations and on future steps.

1 Introduction

Defining conceptual representations along with their associated reasoning procedures required and still requires truly interdisciplinary efforts, involving psychologists (Miller and Charles, 1991; Barsalou, 1999; Malt et al., 2015), philosophers (Machery, 2009; Gärdenfors, 2014), neuroscientists (Vigliocco et al., 2014), and computer scientists (Resnik, 1998; Agirre et al., 2009; Pilehvar and Navigli, 2015). Today, the ever-growing number of applications of semantic technologies demand for further investigation on concepts' meaning: this fact explains the popularity of issues rooted in and related to conceptual similarity, and the success of the present Semantic Word Similarity task (Camacho-Collados et al., 2017).

In this paper we present an approach to the computation of conceptual similarity based on a novel lexical resource, the TTCS^ε—so dubbed after Terms to Conceptual Spaces-Extended—that has been acquired by integrating two different sorts of linguistic resources, such as the encyclopedic knowledge available in BabelNet (Nav-

igli and Ponzetto, 2012) and NASARI (Camacho-Collados et al., 2015), and the common-sense grasped by ConceptNet (Speer and Havasi, 2012). The resulting representation enjoys the interesting property of being anchored to both resources, thereby providing a uniform conceptual access grounded on the sense identifiers provided by BabelNet.

The TTCS^ε provides a conceptual representation inspired to Conceptual Spaces (CSs), a geometric representation framework where knowledge is represented as a set of limited though cognitively relevant quality dimensions (Gärdenfors, 2014). The CSs framework has been recently used to extend and complement the representational and inferential power allowed by formal ontologies with special emphasis on dealing with the corresponding typicality-based conceptual reasoning (Lieto et al., 2015, 2017); in this setting, the TTCS^ε aims at providing a wide-coverage, cognitively based linguistic resource for this sort of knowledge, by extending previous work (Lieto et al., 2016; Mensa et al., 2017).

2 Concept Representation in the TTCS^ε

Concepts representation in the TTCS^ε is consistent with CSs: each concept c is provided with a vector representation \vec{c} providing information on the given concept along some semantic dimensions d . All concepts included in such description are referred to through BabelNet synset IDs, and dimensions themselves are a subset of the relationships available in ConceptNet. Such relations report common-sense information like, e.g., ISA, ATLOCATION, USED FOR, PART OF, MADE OF, HAS A, CAPABLE OF, etc.. For a full description of the employed properties we refer the reader to (Mensa et al., 2017).

Let D be the set of N dimensions. Each con-

<i>RelatedTo</i>	<i>tool, food, table, cutlery, eating, knife, spoon, utensil</i>
<i>IsA</i>	<i>tool, cutlery, utensil</i>
<i>AtLocation</i>	<i>table, desk, plate</i>
<i>UsedFor</i>	<i>eating</i>
<i>Antonym</i>	<i>knife, spoon</i>
<i>MadeOf</i>	<i>metal</i>

Figure 1: Example of representation for the concept FORK (BN:00035902N). The representation has been made human-readable by displaying concept lexicalizations in place of their actual Babel synset IDs.

cept c_i in the linguistic resource is defined as a vector $\vec{c}_i = [s_1^i, \dots, s_N^i]$, where each s_h^i constitutes the set of concepts filling the dimension $d_h \in D$. Each s can contain an arbitrary number of values, or be empty. For example, the representation for the concept FORK includes information about 6 dimensions that are filled with overall 18 concepts, like illustrated in Figure 1.

The TTCS $^{\mathcal{E}}$ resource contains 14,677 concepts, and it was built by starting from the 10K most frequent nouns present in the Corpus of Contemporary American English (COCA),¹ browsing over 11M associations available in ConceptNet and the 2.8M NASARI vectors. Concepts in the TTCS $^{\mathcal{E}}$ are filled, on average, with 14.90 (concept) values.²

3 Conceptual similarity with the TTCS $^{\mathcal{E}}$

Our similarity metrics does not employ WordNet taxonomy and distances between pairs of nodes, such as in (Wu and Palmer, 1994; Leacock et al., 1998), nor it depends on information content accounts either, such as in (Resnik, 1998).

Conversely, given the aforementioned representation for concepts, one principal assumption underlying our approach is that two concepts are

¹<http://corpus.byu.edu/full-text/>.

²The final resource is available for download at the URL <http://ttcs.di.unito.it>.

similar insofar as they share values on the same dimension, such as when they are both used for the same ends, they share the same components, *etc.*. Given two concepts \vec{c}_i and \vec{c}_j , the conceptual similarity along each dimension —filled in both vectors— should be ideally computed as a function of the cardinality of the intersection between overlapping dimensions

$$\text{sim}(\vec{c}_i, \vec{c}_j) = \sum_{k=1}^N |s_k^i \cap s_k^j|$$

where s_k^i is the set of concepts filling the k -th dimension in the vector \vec{c}_i representing the concept c_i . The rationale underlying this formula is to grasp shared features, thereby allowing us to provide an explanation based on common-sense accounts. For example, rather than computing a distance on WordNet or observing how frequently they co-occur, to justify the similarity score for the pair $\langle \textit{bird}, \textit{cock} \rangle$ we consider that each concept ISA ‘animal’; and that both of them are RELATEDTO ‘feather’, ‘chicken’, ‘roosting’ and ‘vertebrate’.

However, our approach is presently limited by the actual average filling factor, and by the noise that can be possibly collected by an automatic procedure built on top of the BabelNet and ConceptNet resources. To handle the possibly unbalanced number of concepts that characterize the different dimensions and to prevent the computation from being biased by more richly defined concepts, we adopt the Symmetrical Tversky’s Ratio Model (Jimenez et al., 2013).

$$\text{sim}(\vec{c}_i, \vec{c}_j) = \frac{1}{N^*} \cdot \sum_{k=1}^{N^*} \frac{|s_k^i \cap s_k^j|}{\beta (\alpha a + (1 - \alpha) b) + |s_k^i \cap s_k^j|}$$

where $|s_k^i \cap s_k^j|$ counts the number of shared concepts that are used as fillers for the dimension d_k in the concept \vec{c}_i and \vec{c}_j , respectively; and a and b are computed as $a = \min(|s_k^i - s_k^j|, |s_k^j - s_k^i|)$, $b = \max(|s_k^i - s_k^j|, |s_k^j - s_k^i|)$; and N^* counts the dimensions actually filled with at least one concept in both vectors. The Symmetrical Tversky’s Ratio Model allows us to tune the balance between cardinality differences (through the parameter α), and between $|s_k^i \cap s_k^j|$ and $|s_k^i - s_k^j|, |s_k^j - s_k^i|$ (through the parameter β). The parameters α and β were set to .8 and .2 for the experimentation, based on a parameter tuning performed on the RG, MC and WS-sim datasets (Rubenstein and Goodenough, 1965;

Dimension (h)	Dim. score	$ s_h^{bird} $	$ s_h^{cock} $	shared values
RELATEDTO	0.68	44	06	<i>feather, chicken roosting, vertebrate</i>
ISA	0.58	07	04	<i>animal</i>

Table 1: Example of computation of the conceptual similarity for the pair ‘bird’ and ‘cock’, by inspecting the actual content of the TTCS^E resource.

Miller and Charles, 1991; Agirre et al., 2009).

For example, in order to compute the semantic similarity between the concepts *bird* and *cock* the TTCS^E finds that the ISA and RELATEDTO dimensions are filled in both \vec{c}_{bird} and \vec{c}_{cock} , thus $N^* = 2$ in the present setting. Figures in column $|s_h^{bird}|$ and $|s_h^{cock}|$ report about how many concepts were retrieved that fill each dimension; elements in the ‘shared values’ column detail how many concepts were found in common to be part of both concept descriptions along the given dimension. The final similarity score obtained by the TTCS^E is 0.63, against the 0.65 assigned in the gold standard.

4 Evaluation

The dataset proposed for the experimentation included 500 word pairs; thanks to the mixture of abstract/concrete concepts and named entities it can be considered as a very complete and challenging test bed. Results have been computed through Pearson and Spearman correlations (respectively, r and ρ) and their harmonic mean; the latter measure ranges between 0.789 (obtained by the LUMINOSO team) and 0, as displayed in Figure 2. In particular, MERALI obtained 0.589 (r), 0.600 (ρ) and 0.594, respectively. We presently focus on this run of the system and disregard the other one that attained substantially similar results, stemming from a slightly different parameters setting.

We dissected the dataset, to identify our system’s weaknesses, to the ends of improving both the conceptual similarity computation procedures and the lexical resource. We noticed that out of the 500 overall word pairs, 405 involve concept comparisons, while in the reminder pairs we have at least one entity at stake (namely, 45 entity-entity pairs and 50 entity-concept pairs).

Comparisons involving entities are somehow different from those involving only concepts: for example, the cases where the semantic similarity is computed between a concept and an entity (e.g., in ‘Darwin-evolution’, ‘Gauss-scientist’,

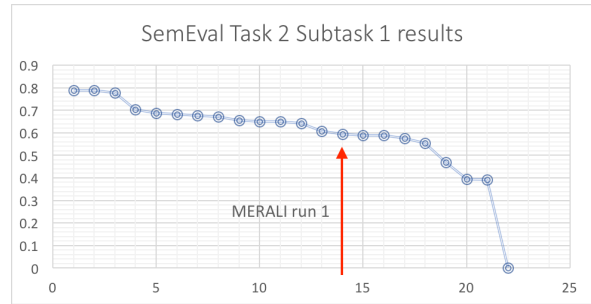


Figure 2: Results of the SemEval Task 2 Subtask 1 (English): harmonic mean between Spearman and Pearson correlations for each team.

‘Siemens-electric train’) pose additional problems with respect to cases in which two entities are considered (such as for ‘Juventus-Bayern Munich’, ‘Plato-Aristotle’, and ‘Alexander Fleming-Penicillin’). Under an ontological perspective, individual entities act like instances, whilst concepts can be considered as classes: one thus wonders what does comparing individuals and classes mean. Moreover, according to the Conceptual Spaces framework, individuals can be thought of as points, while concepts are represented as regions over the multidimensional conceptual space. Comparisons between a class and an individual are intuitively harder in that they require *i*) to find the relations relating the individual and the class being examined; and *ii*) in a CSs perspective, to compare a point with a region. Furthermore, under a cognitive perspective, it is difficult to follow the strategy adopted by humans in providing a score for pairs such as ‘Zara-leggings’ (gold standard similarity judgement: 1.67): directly comparing a manufacturer and a product is nearly unfeasible, since their features can be hardly compared. Justifying the answer is perhaps helpful to give some information on the argumentative paths that can be followed to assess semantic similarity. One major risk, in these respects, is that instead of *similarity*, the scores provided by human annotators rather refer to generic *relatedness*. For example, let us consider the pair ‘tail-Boeing 747’ (gold standard similarity judgement: 1.92): although each Boeing 747 has a tail, the whole plane (holonym) cannot be conceptually similar to its tail (meronym), in the same way a door is not similar to its knob.

So we have re-run the statistical tests to compute Spearman and Pearson correlations over the three subsets (entity-entity, entity-concept, concept-concept); the partial results are reported

	# pairs	r	ρ	harm.mean
entire data	500	0.59	0.60	0.59
entity-entity	45	0.69	0.70	0.69
entity-concept	50	0.72	0.66	0.69
concept-concept	405	0.56	0.57	0.56

Table 2: Spearman (ρ) and Pearson (r) correlations (and their harmonic mean) obtained by the MERALI system over the three subsets.

in Table 2.

It turned out that, against our intuition, the MERALI system has better accuracy on word pairs including an entity; so we further examined the latter subset (concept-concept), where we obtained poorer results. Here we notice that in many cases (22, that is over 5% of this subset) overly high scores were determined by the maximization implemented in the word-similarity: in such cases, in fact, semantic similarity is usually computed as the similarity of the closest senses underlying the given terms (Budanitsky and Hirst, 2006). An example of this sort of errors is the pair ‘apocalypse-fire’ (gold standard similarity judgement: 1.25), where the MERALI system returned a value by far higher than the expected value (namely, 3.85): fires can legitimately be interpreted as apocalyptic events, but only in a figurative way. Similar, though distinct, differences in score are observed when comparing two identical concepts: not always human beings provide the maximum (equality) score, sometimes in unexpected way like for, e.g., ‘movie-film’ (gold standard similarity judgement: 3.92), ‘multiple sclerosis-MS’ (gold standard similarity judgement: 3.92). Out of 24 such cases, for 13 pairs (3% of this subset) we over-estimated the semantic similarity. As regards as fully different concept pairs (46, over 11%), in half cases we have over-estimated the similarity, perhaps due to a too permissive enriching routine that sometimes accepts noisy concepts as dimension fillers.

However, the main issue of the first version of the MERALI system is that the overall amount of information available to the system is often not enough to fully assess the semantic similarity between concepts. Sometimes concepts themselves have been missing, and missing concepts may be lacking in (at least one of) the resources upon which the TTCS^E is built. Also, difficulties stemmed from insufficient information for the

concepts at stake: this phenomenon was observed, e.g., when both concepts have been found, but no common dimension has been filled. This sort of difficulty shows that the coverage of the resource still needs to be enhanced, especially by improving the extraction phase, so to add further concepts *per* dimension, and to fill more dimensions.

5 Conclusions

We have illustrated the system MERALI, that relies on a novel resource, the TTCS^E. The underlying representation is compatible with the Conceptual Spaces framework and aims at putting together encyclopedic and common-sense knowledge. The results of the MERALI system have been illustrated and discussed. The experimentation clearly showed that there is room for improving the system along two main axes: dimensions must be filled with further information, and the quality of the extracted information should be improved. Also the computation of the similarity can be refined by testing further heuristics, so to reduce the cases of over-estimation of semantic similarity. All mentioned aspects will be addressed in our future work.

References

- Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Paşca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of NAACL*. Association for Computational Linguistics, NAACL ’09, pages 19–27.
- Lawrence W Barsalou. 1999. Perceptions of perceptual symbols. *Behavioral and brain sciences* 22(04):637–660.
- Alexander Budanitsky and Graeme Hirst. 2006. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics* 32(1):13–47.
- Jose Camacho-Collados, Mohammad Taher Pilehvar, Nigel Collier, and Roberto Navigli. 2017. Semeval-2017 task 2: Multilingual and cross-lingual semantic word similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017)*. Vancouver, Canada.
- José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. NASARI: a novel approach to a semantically-aware representation of items. In *Proceedings of NAACL*. pages 567–577.
- Peter Gärdenfors. 2014. *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.

- Sergio Jimenez, Claudia Becerra, Alexander Gelbukh, Av Juan Dios Batiz, and Av Mendizabal. 2013. Softcardinality-core: Improving text overlap with distributional measures for semantic textual similarity. In *Proceedings of *SEM 2013*. volume 1, pages 194–201.
- Claudia Leacock, George A Miller, and Martin Chodorow. 1998. Using corpus statistics and word-net relations for sense identification. *Computational Linguistics* 24(1):147–165.
- Antonio Lieto, Enrico Mensa, and Daniele P. Radicioni. 2016. A Resource-Driven Approach for Anchoring Linguistic Resources to Conceptual Spaces. In *Procs of the XV International Conference of the Italian Association for Artificial Intelligence*. Springer, volume 10037 of *LNAI*, pages 435–449. <https://doi.org/10.1007/978-3-319-49130-1>.
- Antonio Lieto, Andrea Minieri, Alberto Piana, and Daniele P. Radicioni. 2015. A knowledge-based system for prototypical reasoning. *Connection Science* 27(2):137–152.
- Antonio Lieto, Daniele P. Radicioni, and Valentina Rho. 2017. Dual PECCS: A Cognitive System for Conceptual Representation and Categorization. *Journal of Experimental & Theoretical Artificial Intelligence* 29(2):433–452. <https://doi.org/10.1080/0952813X.2016.1198934>.
- Edouard Machery. 2009. *Doing without concepts*. Oxford University Press.
- Barbara C Malt, Silvia P Gennari, Mutsumi Imai, Eef Ameen, Noburo Saji, and Asifa Majid. 2015. 11 where are the concepts? what words can and cant reveal. *The conceptual mind: New directions in the study of concepts* page 291.
- Enrico Mensa, Daniele P. Radicioni, and Antonio Lieto. 2017. TTCS²: a Vectorial Resource for Computing Conceptual Similarity. In *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*. Association for Computational Linguistics, Valencia, Spain, pages 96–101. <http://www.aclweb.org/anthology/W17-1912>.
- George A Miller and Walter G Charles. 1991. Contextual correlates of semantic similarity. *Language and cognitive processes* 6(1):1–28.
- Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.* 193:217–250.
- Mohammad Taher Pilehvar and Roberto Navigli. 2015. From senses to texts: An all-in-one graph-based approach for measuring semantic similarity. *Artif. Intell.* 228:95–128.
- Philip Resnik. 1998. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* 11(1).
- Herbert Rubenstein and John B Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8(10):627–633.
- Robert Speer and Catherine Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. In *LREC*. pages 3679–3686.
- Gabriella Vigliocco, Stavroula-Thaleia Kousta, Pasquale Anthony Della Rosa, David P Vinson, Marco Tettamanti, Joseph T Devlin, and Stefano F Cappa. 2014. The neural representation of abstract words: the role of emotion. *Cerebral Cortex* 24(7):1767–1777.
- Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. ACL, pages 133–138.