CROWDSOURCING COMPLEX ASSOCIATIONS AMONG WORDS BY MEANS OF A GAME

Pavel Smrz

Brno University of Technology, Faculty of Information Technology
Bozetechova 2, 61266 Brno, Czech Republic

ABSTRACT

This paper discusses a new approach to creating semantic resources consisting of complex associations among words that can be used for evaluating the content of word embeddings as well as in various language-learning scenarios. We briefly introduce Codenames – an existing party board game – and the way of recording word associations suggested by human players. Advanced word embedding models are then compared on the collected data and it is demonstrated that they often fail in the cases of complex word associations that go beyond simple contextual interchangeability. We conclude with an initial evaluation of the automatic guessing of associated words based on clues provided by human players and a discussion on further extensions of the system towards a wide language coverage and explanations of word associations in the language learning context.

KEYWORDS

Natural Language Processing, Word Embedding, Distributional Semantics, Implicit Crowdsourcing, Games with Purpose, Semantic Representation

1. INTRODUCTION

Crowdsourcing has become a standard in the current era of deep learning from massive data as it reduces costs and increases scalability while the quality can be comparable to traditional expert annotation. Crowdsourcing marketplaces such as Amazon Mechanical Turk or CrowdFlower are often used as platforms for various NLP tasks, including co-reference resolution, word sense disambiguation, sentiment analysis, etc. As opposed to the paid micro-tasks, the implicit crowdsourcing embeds tasks into other forms of activities in order to motivate worker participation. It can take a form of piggybacking on user's behaviour data (such as search keyword and click analysis by search engines), or standalone activities, including reCAPTCHA and various Games with a Purpose (GWAP).

A variety of contexts and particular games have been considered for the creation of valuable language resources by implicit crowdsourcing activities. For example, Duolingo can connect their primary data on learning a foreign language with contributing to machine translation applications [1]. Also, games such as Phrase Detectives [2], Wordrobe [3], or Zombilingo [4] bring novel environments and generate relevant resources that can be employed in NLP applications. However, to the best of our knowledge, none of the existing approaches takes advantage of an existing popular game and extracts usable artefacts from the process of real playing.

This paper discusses such an approach. We employ recordings of the party board game Codenames and extract data relevant to evaluation of lexical semantic models. As the word associations used by a player in the game can cover a plethora of complex relations, we compare

Natarajan Meghanathan et al. (Eds) : CSTY, AI, MaVaS, SIGI, FUZZY - 2019
pp. 75-81, 2019. © CS & IT-CSCP 2019 DOI: 10.5121/csit.2019.91407

existing distributional semantic representations on the recorded word associations and show that some kinds of relations are systematically poorly represented in existing models. We also evaluate the overall quality of the implemented computer player in a variant of the Codenames game.

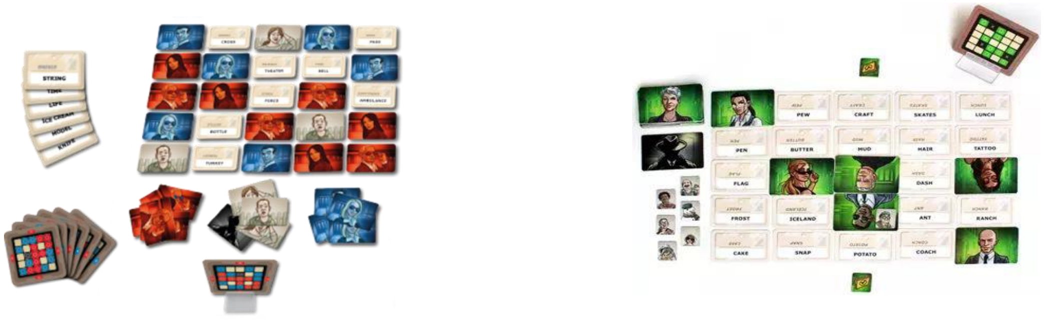
The rest of this paper is organised as follows. The next section briefly introduces the Codenames game and discusses the way existing records are transformed to provide the base of the evaluation. Next, we discuss the type of relations that are typically used and compare them to other relevant language data resources used to evaluate the distribution semantic models. Section 4 then compares results of the most popular word embedding models (Word2Vec, Glove, fastText) and demonstrates results of our automatic player in word guessing. Section 5 summarizes previous research in relevant to our work. Section 6 concludes the paper and outlines directions of our future research work.

2. CODENAMES

Codenames is an original party game for 4 and more players designed by Vlaada Chvátil and published by the Czech Games Edition studio. It has been awarded the prestigious German Spiles des Jahres 2016 and many other relevant awards. Two teams compete by each having a spymaster giving one word clues which can point to multiple words on the board. The other players of the team attempt to guess their team's words while avoiding the words of the other team. As of October 2019, the game has been published in 36 languages.

The 25 words to be in a particular game are randomly chosen from a pool of 400 words and arranged in a 5x5 matrix (see Figure 1). The game plan is also randomly selected, the beginning team has 9 words (agents), the other one 8 words, there are 7 other neutral words (bystanders) and 1 “assassin” word. The clue word should not have the same root as one of the words that are currently in the game and the association needs to be only semantic. Together with the clue word, the spymaster indicates the number of words that are to be associated with the clue (for example, “tree 2” can be a clue for words/agents “nut” and “bark”). The team can then guess up to the number of words given in the current clue and one additional word (usually corresponding to a previous turn clue with a wrong guess). There are also 2 special numbers for the clue. The infinity clue does not suggest an exact number of words to be associated with the current clue word but does not limit the number of guessed words (providing all of them correspond to the words of the guessing team). The zero clue indicates words that the team should avoid in their guessing.

Codenames: Duet is a special variant of the game more suitable for 2 players. Out of the 25 initial words, there are 15 words to be guessed in cooperation by the 2 players. Each player has a separate key with 9 words (out of those 15) to give a clue for. There are 3 “assassin” words for each of the players (the game immediately ends when one of these words is guessed).



a) b)

Figure 1. The Codenames game – (a) Standard and (b) Duet

3. Collected Data Characteristics

The original data set recorded 86 real games (with only human players) with 680 clue-answer pairs. As the data has been additionally intended for the development and evaluation of an automatic game engine, not only the actually guessed words for a clue have been recorded but also the words considered as related during the team discussion. After each game, the data has been further extended by recording the words intended by the spymaster for each particular clue. The graph in Figure 2 shows that the most frequent clues join 2 or 3 words, clues associated with 5 or more words are extremely rare in the dataset.

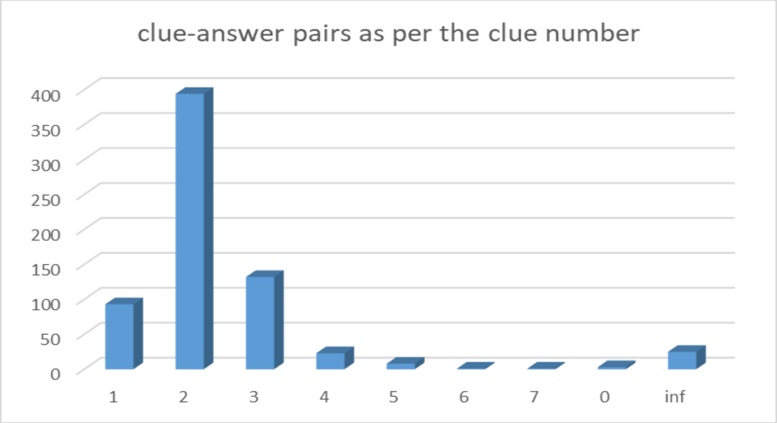


Figure 2. Clue number frequency in the collected data

Recent interest in distributional semantic models opened discussion on the similarity between human knowledge about word meanings and relations and the information represented by the models. Some works such as [5] demonstrate that the distributional representations of words extracted from large corpora do not necessarily well cover typical lexical relations of hypernymy or entailment.

The data collected from the Codenames games indicate that there is a whole bunch of relations that can contribute to the association between a clue word and a set of words in the game. This is particularly true in cases when the spymaster associates more than 2 words with a clue. For example, one has to consider a wide range of relation types to that the clue "tree 4" indicates intended words: *squirrel*, *garden*, *axe*, and *greenhouse*.

4. Experimental Evaluation

Four different models as well as their combinations were evaluated. The most popular word embedding models were considered – Word2Vec [6], Glove [7], and FastText [8], as well as a simple co-occurrence count measure – the Normalized Pointwise Mutual Information (NPMI) given as:

$$NPMI(x, y) = \frac{\log \frac{\mathrm{P}(\mathrm{x} \wedge \mathrm{y})}{\mathrm{P}(\mathrm{x}) \times \mathrm{P}(\mathrm{y})}}{-\log \mathrm{P}(\mathrm{x} \wedge \mathrm{y})} \approx \frac{\log \frac{A \times N}{(A+C) \times (A+B)}}{-\log \frac{A}{N}},$$

where A is the number of co-occurrences of the two words, B is the number of occurrences of the first word without the second one, C is the number of occurrences of the second word without the first one, and N is the total number of contexts in a corpus.

As shown in [9], the most efficient distributed semantic model showed to be fastText. Table 1 compares its accuracy with two other models.

Table 1. Comparison of word embedding models

model	**accuracy**
fastText	74.14%
Word2vec	71.87%
GloVe	68.76%

Consequently, the fastText model was employed, compared, and combined with NPMI. As Table 2 suggests, the OOV approximation helps the standard fastText model. The NPMI overcomes fastText on the collected data but fails in other cases. A weighted combination provides the best results.

Table 2. Best predicting and counting models and their combination

model	**Accuracy**
fastText + OOV approximation	74.99%
NPMI	76.79%
Optimal combination NPMI & fastText	81.33%

Even the best combination of the counting and predicting models provides results that are far from perfect. There are about 17% of errors that correspond to the clue-answer words cases that human players (native speakers) have no problem to guess (and, potentially, explain the association between the clue word and each of the guessed words – see also below).

5. RELATED WORK

Crowdsourcing techniques for creating valuable resources flourished in the last decade. Existing approaches differ in their design, strategy, platforms used, task execution, quality management, and many other attributes. Various meta-studies have been recently published that identify trends and tendencies in the field and suggest future research questions [10, 11, 12]. Several relevant initiatives integrating crowdsourcing techniques for language resources have been also recently established, including the NIEUW (Novel Incentives and Workflows) project funded by the United States National Science Foundation [13] and the European COST Action CA16105: enetCollect – Large-Scale Information Extraction and Gamification for Crowdsourced Language Learning [14], which also provides an umbrella for the research reported in this paper.

Creating language resources by means of GWAP has also a long tradition. Some projects, such as JeuxDeMots [15], can build on more than 10-years' experience. Lessons learned from such an experience clearly show that games need to be attractive, fun, and interesting to be successful in a long-term perspective [16]. The aspect of interestingness is particularly difficult to preserve if a game needs specific training [17] or it involves (potentially long) free-text entries produced by the user [18]. Text adventure games provide an example of attractive environments that can naturally lead to building useful language resources [19]. The use of Codenames as the platform to collect valuable lexical data is also given by the popularity of the game and its attractiveness for various players across languages.

The complexity of relations in human word association networks and the cognitive lexicon have been mainly discussed in relation to word association norms [20, 21]. Basic types of relations (synonymy, antonymy, hypernymy, meronymy) in the data have been identified and shared tasks on their automatic classification have been run [22]. Various approaches have been also proposed for learning vectors that capture the association relationship between words [23, 24, 25]. The discussed need for an explanation of word similarity as well as differences between related words links the work to the research on discriminative attribute identification [26] and explanation mechanisms in natural language inference systems [27].

6. Conclusions and Future Directions

The paper presented a novel approach to create valuable lexical semantic resources by means of playing a popular party game Codenames or its 2-player variant Codenames: Duet. It also discussed performance of existing word embedding models and a co-occurrence counting method on the collected data. Although the weighted combination of the mentioned techniques enables correctly guessing the suggested words in more than 80% of the recorded cases, there is still a significant room for improvement. An implementation of the Duet game in English can be tested at http://athena3.fit.vutbr.cz:8086/duet_init_en.

Our future work will employ the concept of the Codenames game in language learning. As a part of our activity in the European COST Action CA16105: enetCollect – Large-Scale Information Extraction and Gamification for Crowdsourced Language Learning – we will organize a hackathon focusing on the selection of appropriate words corresponding to particular learning needs, transferring language pre-processing steps across languages (identification of the same roots of words, suggesting hypernyms, updating the underlying models to cover recent news and/or domain-specific knowledge, etc.). We will also further extend our cooperation with the game publisher with the ultimate goal to implement a computer companion for the game across languages.

In a longer perspective, we will also explore the potential of the Codenames game as a means to encourage lexicon exploration in computer-assisted language learning. The experience from the game plays by non-native speakers indicates that some associations among words are language-or culture-specific and that the discussion on the reasons why particular words are related brings valuable insights for language learners and piques their curiosity. To be able to explain the word associations suggested by a word embedding model, we will employ standard word embedding visualisation approaches [28] as well as tools such as the TensorFlow Projector[1] or EmbVis[2]. To explore the similar contexts the words appear in, we will adapt the idea of word sketches [29].

Acknowledgements

This work was supported by the Czech Ministry of Education, Youth and Sports, subprogram INTERCOST, project code: LTC18006 – Large-Scale Information Extraction and Gamification for Crowdsourced Language Learning.

References

[1] Von Ahn, Luis. Duolingo: learn a language for free while helping to translate the web. In Proceedings of the 2013 international conference on Intelligent user interfaces, pp. 1-2. ACM, 2013.

[2] Poesio, Massimo, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo, and Luca Ducceschi. Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation. ACM Transactions on Interactive Intelligent Systems (TiiS) 3, no. 1 (2013): 3.

[3] Jurgens, David, and Roberto Navigli. It’s all fun and games until someone annotates: Video games with a purpose for linguistic annotation. Transactions of the Association for Computational Linguistics 2 (2014): 449-464.

[4] Fort, Karën, Bruno Guillaume, and Hadrien Chastant. Creating Zombilingo, a Game With A Purpose for dependency syntax annotation. 2014.

[5] Levy, Omer, Steffen Remus, Chris Biemann, and Ido Dagan. Do supervised distributional methods really learn lexical inference relations? In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 970-976. 2015.

[6] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111-3119. 2013.

[7] Pennington, Jeffrey, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.

[8] Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651 (2016).

[9] Jares, Petr. Computer as an Intelligent Partner in the Word-Association Game Codenames. Bachelor Thesis. Brno University of Technology, 2019.

[10] Neto, Fábio R. Assis, and Celso AS Santos. Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management. Information Processing & Management 54, no. 4 (2018): 490-506.

[11] Ghezzi, Antonio, Donata Gabelloni, Antonella Martini, and Angelo Natalicchio. Crowdsourcing: a review and suggestions for future research. International Journal of Management Reviews 20, no. 2 (2018): 343-363.

[12] Qarout, Rehab, Alessandro Checco, Gianluca Demartini, and Kalina Bontcheva. Platform-Related Factors in Repeatability and Reproducibility of Crowdsourcing Tasks. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, vol. 7, no. 1, pp. 135-143. 2019.

[13] Cieri, Christopher, James Fiumara, Mark Liberman, Chris Callison-Burch, and Jonathan Wright. Introducing NIEUW: Novel incentives and workflows for eliciting linguistic data. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018.

[14] Lyding, Verena, Lionel Nicolas, Branislav Bédi, and Karën Fort. Introducing the European NETwork for Combining Language LEarning and Crowdsourcing Techniques (enetCollect). Future-proof CALL: language learning as exploration and encounters–short papers from EUROCALL 2018 (2018): 176.

[15] Lafourcade, Mathieu. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In SNLP'07: 7th international symposium on natural language processing, p. 7. 2007.

[16] Joubert, Alain, Mathieu Lafourcade, and Nathalie Le Brun. The JeuxDeMots Project is 10 Years Old: What We Have Learned. Games and Gamification for Natural Language Processing (2018): 22.

[17] Habernal, Ivan, Raffael Hannemann, Christian Pollak, Christopher Klamm, Patrick Pauli, and Iryna Gurevych. Argotario: Computational argumentation meets serious games. arXiv preprint arXiv:1707.06002 (2017).

[18] Lafourcade, Mathieu, and Alain Joubert. TOTAKI: A help for lexical access on the TOT problem. In Language Production, Cognition, and the Lexicon, pp. 95-112. Springer, Cham, 2015.

[19] Urbanek, Jack, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, and Jason Weston. Learning to speak and act in a fantasy text adventure game. arXiv preprint arXiv:1903.03094 (2019).

[20] Lubaszewski, Wiesław, Izabela Gatkowska, and Maciej Godny. How a Word of a Text Selects the Related Words in a Human Association Network. In Cognitive Approach to Natural Language Processing, pp. 41-62. Elsevier, 2017.

[21] Gaume, Bruno, Lydia Mai Ho-Dac, Ludovic Tanguy, Cécile Fabre, Bénédicte Pierrejean, Nabil Hathout, Jérôme Farinas et al. Toward a Computational Multidimensional Lexical Similarity Measure for Modeling Word Association Tasks in Psycholinguistics. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pp. 71-76. 2019.

[22] Santus, Enrico, Anna Gladkova, Stefan Evert, and Alessandro Lenci. The CogALex-V shared task on the corpus-based identification of semantic relations. In Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V), pp. 69-79. 2016.

[23] Bel-Enguix, Gemma, Helena Gómez-Adorno, Jorge Reyes-Magaña, and Gerardo Sierra. Wan2Vec: Embeddings learned on word association norms. Semantic Web Preprint (2019):1-16.

[24] Joshi, Mandar, Eunsol Choi, Omer Levy, Daniel S. Weld, and Luke Zettlemoyer. pair2vec: Compositional word-pair embeddings for cross-sentence inference. arXiv preprint arXiv:1810.08854 (2018).

[25] Camacho-Collados, Jose, Luis Espinosa-Anke, Shoaib Jameel, and Steven Schockaert. A Latent Variable Model for Learning Distributional Relation Vectors. In International Joint Conferences on Artificial Intelligence. 2019.

[26] Krebs, Alicia, Alessandro Lenci, and Denis Paperno. Semeval-2018 task 10: Capturing discriminative attributes. In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 732-740. 2018.

[27] Thorne, James, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. Generating Token-Level Explanations for Natural Language Inference. arXiv preprint arXiv:1904.10717 (2019).

[28] Heimerl, Florian, and Michael Gleicher. Interactive analysis of word vector embeddings. In Computer Graphics Forum, vol. 37, no. 3, pp. 253-265. 2018.

[29] Kilgarriff, A., Rychly, P., Smrz, P. and Tugwell, D., 2004. Itri-04-08 The Sketch engine. Information Technology, 105, p.116.