

Scaling up Analogy with Crowdsourcing and Machine Learning

Joel Chan¹, Tom Hope², Dafna Shahaf² and Aniket Kittur¹

¹ Human-Computer Interaction Institute
Carnegie Mellon University, Pittsburgh PA 15213, USA
joelchuc@cs.cmu.edu, nkittur@cs.cmu.edu,
² School of Computer Science and Engineering
Hebrew University of Jerusalem, Jerusalem, Israel
tom.hope@mail.huji.ac.il, dshahaf@cs.huji.ac.il

Abstract. Despite tremendous advances in computational models of human analogy, a persistent challenge has been scaling up to find useful analogies in large, messy, real-world data. The availability of large idea repositories (e.g., the U.S. patent database) could significantly accelerate innovation and discovery in a way never previously possible. Previous approaches have been limited by relying on hand-created databases that have high relational structure but are very sparse (e.g., predicate calculus representations). Traditional machine-learning/information-retrieval similarity metrics (e.g., LSA) can scale to large, natural-language datasets; however, while these methods are good at detecting surface similarity, they struggle to account for structural similarity. In this paper, we propose to leverage crowdsourcing techniques to construct a dataset with rich “analogy-tuning” signals, used to guide machine learning models towards matches based on relations rather than surface features. We demonstrate our approach with a crowdsourced analogy identification task, whose results are used to train deep learning algorithms. Our initial results suggest that a deep learning model trained on positive/negative example analogies from the task can find more analogous matches than an LSA baseline, and that incorporating behavioral signals (such as queries used to retrieve an analogy) can further boost its performance.

Keywords: Analogy, crowdsourcing, machine learning

1 Introduction

Invention by analogy (i.e., transferring ideas from other domains that are structurally similar to a target problem) is a powerful way to create new innovations. For example, a car mechanic invented a new low-cost way to ease difficult childbirth by drawing an analogy to a cork extraction method in wineries (inserting and inflating a small plastic bag in the bottle) [12]. This award-winning device has the potential to change lives worldwide, particularly women in developing countries with limited medical resources.

The recent growth of online innovation repositories represents an unparalleled opportunity for invention by analogy. These repositories contain hundreds