cases

doc_1		doc_2		decision	id
authors	• Iz Beltagy • Stephen Roller • Pengxiang Cheng • K. Erk • R. Mooney	**authors**	• Iz Beltagy • Stephen Roller • Pengxiang Cheng • K. Erk • R. Mooney	NOT DUPLICATES	447
title	Representing Meaning with a Combination of Logical Form and Vectors	**title**	Representing Meaning with a Combination of Logical and Distributional Models		
publication_date	2015-05-26 00:00:00	**publication_date**	2015-05-26 00:00:00		
source	SupportedSources.SEMANTIC_SCHOLAR	**source**	SupportedSources.SEMANTIC_SCHOLAR		
journal	ArXiv	**journal**	Computational Linguistics		
volume	abs/1505.06816	**volume**	42		
doi		**doi**	10.1162/COLI_a_00266		
urls	• https://www.semanticscholar.org/paper/fca1e631b8f93036065311eb92727c509423475a	**urls**	• https://www.semanticscholar.org/paper/d799ab952c0f677d4526c485cc14555b7265ecc6		
id	id-7787215108838565366	**id**	id3338067326014052050		
abstract	NLP tasks differ in the semantic information they require, and at this time no single semantic representation fulfills all requirements. Logic-based representations characterize sentence structure, but do not capture the graded aspect of meaning. Distributional models give graded similarity ratings for words and phrases, but do not adequately capture overall sentence structure. So it has been argued that the two are complementary. In this paper, we adopt a hybrid approach that combines logic-based and distributional semantics through probabilistic logic inference in Markov Logic Networks (MLNs). We focus on textual entailment (RTE), a task that can utilize the strengths of both representations. Our system is three components, 1) parsing and task representation, where input RTE problems are represented in probabilistic logic. This is quite different from representing them in standard first-order logic. 2) knowledge base construction in the form of weighted inference rules from different sources like WordNet, paraphrase collections, and lexical and phrasal distributional rules generated on the fly. We use a variant of Robinson resolution to determine the necessary inference rules. More sources can easily be added by mapping them to logical rules; our system learns a resource-specific weight that counteract scaling differences between resources. 3) inference, where we show how to solve the inference problems efficiently. In this paper we focus on the SICK dataset, and we achieve a state-of-the-art result. Our system handles overall sentence structure and phenomena like negation in the logic, then uses our Robinson resolution variant to query distributional systems about words and short phrases. Therefor, we use our system to evaluate distributional lexical entailment approaches. We also publish the set of rules queried from the SICK dataset, which can be a good resource to evaluate them.	**abstract**	NLP tasks differ in the semantic information they require, and at this time no single semantic representation fulfills all requirements. Logic-based representations characterize sentence structure, but do not capture the graded aspect of meaning. Distributional models give graded similarity ratings for words and phrases, but do not capture sentence structure in the same detail as logic-based approaches. It has therefore been argued that the two are complementary.We adopt a hybrid approach that combines logical and distributional semantics using probabilistic logic, specifically Markov Logic Networks. In this article, we focus on the three components of a practical system:1 1) Logical representation focuses on representing the input problems in probabilistic logic; 2) knowledge base construction creates weighted inference rules by integrating distributional information with other sources; and 3) probabilistic inference involves solving the resulting MLN inference problems efficiently. To evaluate our approach, we use the task of textual entailment, which can utilize the strengths of both logic-based and distributional representations. In particular we focus on the SICK data set, where we achieve state-of-the-art results. We also release a lexical entailment data set of 10,213 rules extracted from the SICK data set, which is a valuable resource for evaluating lexical entailment systems.2		
versions		**versions**			