Associative Reasoning for Commonsense Knowledge

Claudia Schon

Institute for Web Science and Technologies, Universität Koblenz, Universitätsstraße 1, 56070 Koblenz, Germany

Abstract

Associative reasoning refers to the human ability to focus on knowledge that is relevant to a particular problem. In this process, the meaning of symbol names plays an important role: when humans focus on relevant knowledge about the symbol *ice*, similar symbols like *snow* also come into focus. In this paper, we model this associative reasoning by introducing a selection strategy that extracts relevant parts from large commonsense knowledge sources. This selection strategy is based on word similarities from word embeddings and is therefore able to take the meaning of symbol names into account. We demonstrate the usefulness of this selection strategy with a case study from creativity testing.

Keywords

selection strategies, commonsense knowledge

1. Introduction

According to Kahneman [1], humans rely on two different systems for reasoning. System 1 is fast, emotional and less accurate, system 2 is slow, more deliberate and logical. Reasoning with system 2 is much more difficult and exhausting for humans. Therefore, system 1 is usually used first to solve a task and system 2 is only used for demanding tasks. Examples of tasks that system 1 does are solving simple math problems like 3 + 3, driving a car in an empty street, or associating a certain profession with the description like *a quiet, shy person who prefers to deal with numbers rather than people*. Especially, the associative linking of information with each other falls within the scope of system 1. In contrast, we typically use system 2 for tasks that require our full concentration such as driving in a crowded downtown area or following complex logical reasoning.

Humans have vast amounts of background knowledge that they skillfully use in reasoning. In doing so, they are able to focus on knowledge that is relevant for a specific problem. Associative thinking and priming play an important role in this process. These are things handled by system 1. The human ability of focusing on relevant knowledge is strongly dependent on the meaning of symbol names. When people focus on relevant background knowledge for a statement like *The pond froze over for the winter.*, similarities of symbols play an important role. For this statement, a human will certainly not only focus on background knowledge that relates exactly to the terms *pond*, *froze*, and *winter*, but also knowledge about similar term such as *ice* and *snow*. We refer to the process of focusing on relevant knowledge as *associative reasoning*.

schon@uni-koblenz.de (C. Schon)
https://userpages.uni-koblenz.de/~obermaie/ (C. Schon)

If we want to model the versatility of human reasoning, it is necessary to model not only different types of reasoning such as deductive, abductive, and inductive reasoning, but also the ability to focus on relevant background knowledge using associative reasoning.

This aspect of human reasoning, focusing on knowledge relevant to a problem or situation, is what we model in this paper. For this purpose, we develop a selection strategy that extracts relevant parts from a large knowledge base containing background knowledge. We use background knowledge formalized in knowledge graphs like ConceptNet [2], ontologies like Adimen SUMO [3], and Cyc [4], or knowledge bases. A nice property of commonsense knowledge sources is that the used symbol names are often based on natural language words. For example in Adimen SUMO you find symbol names like *c__SecondarySchool*. To model the associative nature of human focusing, we exploit this nice property and use word similarities from word embeddings.

In word embeddings, large amounts of text are used to learn a vector representation of words. These so called word vectors have the nice property that similar words are represented by similar vectors. We propose a representation of background knowledge in terms of vectors such that similar statements in the background knowledge are represented by similar vectors.

Based on this vector representation of background knowledge, we present a new selection strategy, the *vector-based selection* that pays attention to the meaning of symbol names and thus models associative reasoning as it is done by humans. The main contributions of this paper are:

- The introduction of the vector-based selection strategy, a statistical selection technique for commonsense knowledge which is based on word embeddings.
- A case study using benchmarks for creativity testing in humans which demonstrates that the vector-based selection allows to model associative reasoning and selects commonsense knowledge in a very focused way.

The paper is structured as follows: after discussing related work in Sec. 2 and preliminaries in Sec. 3, we briefly revise SInE, a selection strategy for first-order logic reasoning with large theories in Sec. 4. Next, we turn to the integration of statistical information into selection strategies in Sec. 5 where, after revising distributional semantics , we introduce the vector-based selection strategy. In Sec. 6 we present experimental results. Finally, we discuss future work.

2. Related Work

Selecting knowledge that is relevant to a specific problem is also an important task in automatic theorem proving. In this area, often a large set of axioms called a knowledge base is given as background knowledge, together with a much smaller set of axioms $F_1, \dots, F_n$ and a query Q. The reasoning task of interest is to show that the knowledge base together with the axioms $F_1, \dots, F_n$ implies the query Q. This corresponds to showing that $F_1 \wedge \dots \wedge F_n \rightarrow Q$ is entailed by the knowledge base. $F_1 \wedge \dots \wedge F_n \rightarrow Q$ is usually referred to as *goal*. As soon as the size of the knowledge base forbids to use the entire knowledge base to show that Q follows from the knowledge base using an automated theorem prover, it is necessary to select the axioms from the knowledge base that are necessary for this reasoning task. However, identifying these axioms is not trivial, so common selection strategies are based on heuristics and are usually incomplete. This means that it is not always possible to solve the reasoning task with the

The pond froze over for the winter. What happened as a result?

1. People brought boats to the pond.
2. People skated on the pond.

Figure 1: Example from the *Choice of Plausible Alternative Challenge* (COPA) [13].

selected axioms: If too few axioms have been selected, the prover cannot find a proof. If too many have been selected, the reasoner may be overwhelmed with the set of axioms and run into a timeout.

Most strategies for axiom selection are purely syntactic like the SInE selection [5], lightweight relevance filtering [6] and axiom relevance ordering [7]. A semantic strategy for axiom selection is SRASS [8] which is a model-based approach. This strategy is based on the computation of models for subsets of the axioms and consecutively extends these sets. Another interesting direction of research is the development of metrics for the evaluation of selection techniques [9] which allow to measure the quality of selection strategies without having to actually run the automated theorem prover on the selected axioms and the conjecture at hand Another approach to axiom selection is the use of formula metrics [10] which measure the dissimilarity of different formulae and lead to selection strategies which allows to select the k axioms from a knowledge base most similar to a given problem. None of the selection methods mentioned so far in this section take the meaning of symbol names into account.

An area where the meaningfulness of symbol names was evaluated is the semantic web [11]. The authors come to the conclusion that the semantics encoded in the names of IRIs (Internationalized Resource Identifiers) carry a kind of social semantics which coincides with the formal meaning of the denoted resource.

Similarity SInE [12] is an extension of SInE selection which uses a word embedding to take similarity of symbols into account. By this mixture of syntactic and statistical methods, Similarity SInE represents a hybrid selection approach. In contrast, the vector-based selection presented in this paper is a purely statistical approach.

3. Preliminaries and Task Description

Numerous sets of benchmarks exist for the area of commonsense reasoning. Typically these problems are multiple choice questions about everyday situations which are given in natural language. Fig. 1 shows a commonsense reasoning problem from the choice of plausible alternative challenge (COPA) [13]. Usually, for these commonsense reasoning problems it is not the case that one of the answer alternatives can actually be logically inferred. Often only one of the answer alternatives is more plausible than the others. To solve these problems, a broad background knowledge is necessary. For the example given in Fig. 1, knowledge about winter, ice, frozen surfaces and boats is necessary. In humans, system 1 with associative reasoning is responsible to focus on relevant background knowledge for a specific problem.

In this paper, we aim at modeling the human ability to focus on background knowledge relevant for a specific task. We introduce a selection strategy based on word embeddings to achieve this. For this, we assume that the background knowledge is given in first-order logic. One reason for this assumption is that this allows to use already existing automated theorem provers for modeling human reasoning in further steps. Furthermore, this allows to easily compare our approach to selection strategies for first-order logic theorem proving like SInE. Moreover, this assumption is not a limitation, since knowledge given in other formes like for example in the form of a knowledge graph can be easily transformed into first-order logic [14].

We furthermore assume that the description of the commonsense reasoning problem is given as a first-order logic formula. Again, this is not a limitation since, for example, the KnEWS [15] system can convert natural language into first-order logic formulas. Following terminology from first-order logic reasoning, we refer to the formula for the commonsense reasoning problem as *goal.* Referring to the example from Fig. 1, we would denote the first-order logic formula for the statement *The pond froze over for the winter.* as F, the formula for *People brought boats to the pond.* as Q_1, and the formula for *People skated on the pond.* as Q_2. This leads to the two goals $F \rightarrow Q_1$ and $F \rightarrow Q_2$ for the commonsense reasoning problem from Fig. 1. For these goals, we could now select from knowledge bases with background knowledge using first-order logic axiom selection techniques.

Axiom selection for a given goal in first-order logic as described at the beginning of Sect. 2 is very similar to the problem of selecting background knowledge relevant for a specific problem in commonsense reasoning. Both problems have in common that large amounts of background knowledge are given that is too large to be considered completely. The main difference is the fact that in commonsense reasoning we cannot necessarily assume that a proof for a certain goal can be found. Therefore, drawn inferences are also interesting in this domain. In both cases, the task is to select knowledge that is relevant for the given goal formula.

In the case study in Sect. 6, we will compare the vector-based selection strategy presented in Sect. 5 with the syntax-based SInE selection strategy which is broadly used in first-order logic theorem proving. Therefore, we briefly introduce the SInE selection in the next section.

In the following we denote the set of all predicate and function symbols occurring in a formula F by $sym(F)$. We slightly exploit notation and use $sym(KB)$ for the set of all predicate and function symbols occurring in a knowledge base KB.

4. SInE: a Syntax-Based Selection Strategy

In [5] the SInE selection strategy is introduced which is successfully used by many automated theorem provers. Since this selection strategy does not consider the meaning of symbol names, we classify this strategy as a *syntax-based* selection. The basic idea of SInE is to determine a set of symbols for each axiom in the knowledge base which is allowed to *trigger* the selection of this axiom. For this a *trigger* relation is defined as follows:

Definition 4.1 (Trigger relation for the SInE selection [5])**.** Let KB be a knowledge base, A be an axiom in KB and $s \in sym(A)$ be a symbol. Let furthermore $occ(s, KB)$ denote the number of axioms in which s occurs in KB and $t \in \mathbb{R}$, $t \geq 1$. Then the triggers relation is defined as

$$triggers(s, A) \text{ iff for all symbols } s' \text{ occurring in } A \text{ we have } occ(s, KB) \leq t \cdot occ(s', KB) \quad (1)$$

Note that an axiom can only be triggered by symbols occurring in the axiom. Parameter t specifies how strict we are in selecting the symbols that are allowed to trigger an axiom. For $t = 1$ (the default setting of SInE), a symbol s may only trigger an axiom A if there is no symbol s' in A that occurs less frequently in the knowledge base than s. This prevents frequently occurring symbols such as *subClass* and *instanceOf* from being allowed to trigger all axioms they occur in.

The *triggers* relation is then used to select axioms for a given goal. The basic idea is that starting from the symbols occurring in the goal, the symbols occurring in the goal are considered to be relevant and an axiom A is selected if A is triggered by some symbol occurring in the set of relevant symbols. The symbols occurring in the selected axioms are added to the set of relevant symbols and if desired, the selection can be repeated.

Definition 4.2 (Trigger-based selection [5])**.** Let KB be a knowledge base, A be an axiom in KB and $s \in sym(KB)$. Let furthermore G be a goal to be proven from KB.

1. If s is a symbol occurring in the goal G, then s is 0-step triggered.
2. If s is n-step triggered and s triggers A ($triggers(s, A)$), then A is $n + 1$-step triggered.
3. If A is n-step triggered and s occurs in A, then s is n-step triggered, too.

An axiom or a symbol is called triggered if it is n-step triggered for some $n \geq 0$.

For a given knowledge base, goal G and some $n \in \mathbb{N}$ SInE selects all axioms which are n-step triggered. In the following the SInE selection selecting all n-step triggered axioms is called SInE with recursion depth n.

SInE selection can also be used in commonsense reasoning to select background knowledge relevant to a statement: to do this, we just need to convert this statement into a first-order logic formula, and use the formula as a goal and select with SInE for it.

5. Use of Statistical Information for the Selection of Axioms

SInE selection completely ignores the meaning of symbol names. For SInE it makes no difference whether a predicate is called p or *dog*. If we consider knowledge bases with commonsense knowledge, the meaning of symbol names provides information that can be exploited by a selection strategy. For example, the symbol *dog* is more similar to the symbol *puppy* than to the symbol *car*. If a goal containing the symbol *dog* is given, it is more reasonable to select axioms containing the symbol *puppy* than axioms containing the symbol *car*. This corresponds to human associative reasoning, which also takes into account the meaning of symbol names and similarities.

5.1. Distributional Semantics

To determine the semantic similarity of symbol names, we rely on distributional semantics of natural language, which is used in natural language processing. The basic idea of distributional semantics is best explained by a quote from Firth, one of the founders of this approach:

> You shall know a word by the company it keeps. [16]

The basis of distributional semantics is the distributional hypothesis [17], according to which words with similar distributional properties on large texts also have similar meaning. In other words: Words that occur in a similar context are similar.

An approach used in many domains which is based on the distributional hypothesis are word embeddings [18, 19]. Word embeddings map the words of a vocabulary to vectors in $\mathbb{R}^n$. Typically, word embeddings are learned using neural networks on very large text sets. WeSince we use existing word embeddings in the following, we do not go into the details of creating word embeddings. An interesting property of word embeddings is that semantic similarity of words corresponds to the relative similarities of the vector representations of those words. To determine the similarity of two vector representations the cosine similarity is usually used.

Definition 5.1 (Cosine similarity of two vectors)**.** Let $u, v \in \mathbb{R}^n$, both non-zero. The *cosine similarity* of u and v is defined as:

$$cos\_sim(u, v) = \frac{u \cdot v}{||u|| \, ||v||}$$

The cosine similarity of two vectors u and v takes values between -1 and 1. For exactly opposite vectors the value is -1, for orthogonal vectors the value is 0 and for equal vectors the value 1. The more similar two vectors are, the greater is their cosine similarity. For example in the ConceptNet Numberbatch word embedding [2], the cosine similarity of *dog* and *puppy* is 0.84140545 which is much larger than the cosine similarity of *dog* and *car* that is 0.13056317. Based on these similarities, word embeddings can furthermore be used to determine the k words in the vocabulary most similar to a given word for some $k \in \mathbb{N}$.

5.2. Vector-Based Selection: A Statistical Selection Strategy

Word embeddings represent words as vectors in such a way that words that are frequently used in a similar context are mapped to similar vectors. Vector-based selection aims to represent the axioms of a knowledge base as vectors in such a way that similar axioms are mapped to similar vectors. Where we consider two axioms of a knowledge base to be similar if they represent similar knowledge.

Fig. 2 gives an overview of the vector-based selection strategy. In a preprocessing step, vector representations are computed for all axioms of the knowledge base using an existing word embedding. This preprocessing step has to be performed only once. Given a goal G for which we want to check if it is entailed by the knowledge base, we transform G into a vector representation using the same word embedding as for the vector transformation of the knowledge base. Next, vector-based selection determines the k vectors in the vector representation of the knowledge base most similar to the vector representation of goal G. The corresponding k axioms form the result of the selection. Various metrics can be used for determining the k vectors that are most similar to the vector representation of G. We use cosine similarity, which is also widely used in word embeddings.

One way to represent an axiom as a vector is to look up the vectors of all the symbols occurring in the axiom in the word embedding and represent the axiom by the average of these vectors. However this treats all symbols occurring in an axiom equally. This is not always useful, as the axiom in Fig. 3 from Adimen SUMO illustrates for which it seems desirable that the symbols

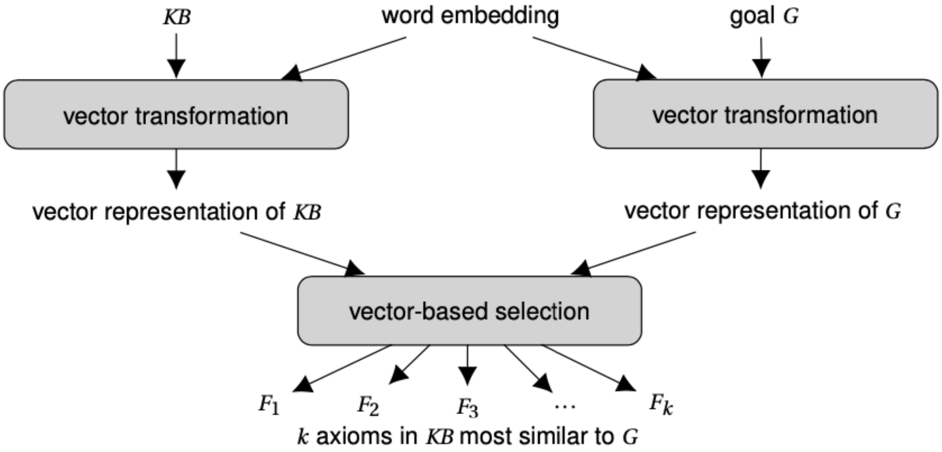

Figure 2: Overview of the vector-based selection strategy. The vector transformation of the knowledge base *KB* and the vector transformation of the goal use the same word embedding.

instance, *agent* and *patient* contribute less to the computation of the vector representation than the symbols *carnivore*, *eating* and *animal*. The reason for this lies in the frequency of the symbols in the knowledge base which are given in the Table in Fig. 3. Symbols *carnivore*, *eating* and *animal* occur much less frequently in Adimen SUMO than *instance*, *agent* and *patient*. This suggests that *carnivore*, *eating* and *animal* are more important for the statement of the axiom. This is similar to the idea in SInE that only the least common symbol in an axiom is allowed to trigger the axiom. We implement this idea in the computation of the vector representation of axioms by weighting the influence of a symbol using inverse document frequency (idf). In the area of information retrieval, for the task of rating the importance of word w to a document d in a set of documents D, idf is often used to diminish the weight of a word that occurs very frequently in the set of documents. Assuming that there is at least one document in D, in which w occurs, $idf(w, D)$ is defined as:

$$idf(w, D) = \log \frac{|D|}{|\{d \in D \mid w \text{ occurs in } d\}|}$$

If w occurs in all documents in D, the fraction is equal to 1 and $idf(w, D) = 0$. If w occurs in only one of the documents in D, the fraction is equal to $|D|$ and $idf(w, D) > 0$. The higher the proportion of documents in which w occurs, the lower $idf(w, D)$.

We transfer this idea to knowledge bases by interpreting a knowledge base as a set of documents and each axiom in this knowledge base as a document. The resulting computation of idf for a symbol in a knowledge base is given in Def. 5.2. For the often used tf-idf (term frequency - inverse document frequency) the idf value is multiplied by the term frequency of a term in a certain document. However since the number of occurrences of a symbol in a single axiom does not necessarily correspond to its importance to the axiom (as illustrated by the axiom given in Fig. 3), we omit this multiplication and use idf for the weighting instead. Multiplying the idf value of a symbol with its tf value in a formula could even increase the

$$\forall X, Y, Z\Big(\big(instance(X, carnivore) \wedge instance(Y, eating)$$
$$\wedge\, agent(Y, X) \wedge patient(Y, Z)\big) \rightarrow instance(Z, animal)\Big)$$

Symbol Name:	instance	agent	patient	carnivore	eating	animal
Frequency:	4237	140	183	5	6	63

Figure 3: Example axiom from Adimen SUMO together with frequencies of the symbols of the axiom in Adimen SUMO. To increase readability, we omitted prefixes of symbols.

influence of frequent symbols like *instance*, since they often appear more than once in a formula.

For simplicity, we assume that $sym(F)$ is a subset of the vocabulary of the word embedding in the following definition.

Definition 5.2 (idf-based vector representation of an axiom, a knowlege base)**.** Let $KB = \{F_1, \dots, F_n\}$, $n \in \mathbb{N}$ be a knowledge base, $F \in KB$ be an axiom. V be a vocabulary and $f : V \rightarrow \mathbb{R}^n$ a word embedding. Let furthermore $sym(F) \subseteq V$. The *idf* value for a symbol $s \in sym(F)$ w.r.t. KB is defined as

$$idf(s, KB) = \log \frac{|KB|}{|\{F' \in KB \mid s \in sym(F')\}|}$$

The *idf-based vector representation* of F is defined as

$$v_{idf}(F) = \frac{\sum_{s \in sym(F)} idf(s, KB) \cdot f(s)}{\sum_{s \in sym(F)} idf(s, KB)}$$

Furthermore, $V_{idf}(KB) = \{v_{idf}(F_1), \dots, v_{idf}(F_n)\}$ denotes the idf-based vector representation of KB.

Note that this definition completely ignores the structure of axioms resulting that axiom $\forall X(animal(X) \wedge fluffy(X))$ is represented by the same vector as $\forall X(animal(X) \vee fluffy(X))$. However, this is not a disadvantage, since our goal is a selection of axioms that matches the topic of a goal, and therefore we need the vector representation of an axiom to represent only the topic and not the exact statement of the axiom.

Given a goal G and a knowledge base, we can use the vector representations of the knowledge base and G to select the k axioms from the knowledge base most similar to the vector representation of G for some $k \in \mathbb{N}$ (see Fig. 2).

Definition 5.3 (Vector-based selection)**.** Let KB be a knowledge base, G be a goal with $sym(G) \subseteq sym(KB)$ and $f : V \rightarrow \mathbb{R}^n$ a word embedding. Let furthermore V_{KB} be a vector representation of KB and v_G a vector representation for G both constructed using f. For $k \in \mathbb{N}$, $k \leq |KB|$ the k axioms in KB most similar to G are given as

$$\begin{aligned} mostsimilar(KB, G, k) = \{F_1, \dots, F_k \mid\ & \{F_1, \dots, F_k\} \subseteq KB \text{ and} \\ & \forall F' \in KB \setminus \{F_1, \dots, F_k\} \\ & cos\_sim(v_{F'}, v_G) \leq \min_{i=1,\dots,k} cos\_sim(v_{F_i}, v_G)\}. \end{aligned}$$

For KB, G and $k \in \mathbb{N}$ given as above described, vector-based selection selects $mostsimilar(KB, G, k)$.

Def. 5.3 is intentionally very general and allows other vector representations besides idf-based vector representation. Furthermore, the similarity measure *cos_sim* can be easily replaced by some other measure like euclidean distance.

In the previous section we assumed the set of symbols in a knowledge base to be a subset of the vocabulary of the used word embedding. However in practice this is not always the case and in many cases it might be necessary to construct a mapping for this. Each combination of knowledge base and word embedding requires a specific mapping. As an example we describe in [20] how we generated different mappings to relate the symbols in knowledge base Adimen SUMO [3] to the vocabulary of the ConceptNet Numberbatch word embedding. For the case study we present in the next section such a mapping is not necessary which is why we refrain from presenting it.

6. Evaluation: A Case Study on Commonsense Knowledge

In areas where commonsense knowledge is used as background knowledge, automated theorem provers can be used not only for finding proofs, but also as inference engines. One reason for this is that even if there are large ontologies and knowledge bases with commonsense knowledge, this knowledge is still incomplete. Therefore, it is likely that not all the information needed for a proof is represented. Nevertheless, automated theorem provers can be very helpful on commonsense knowledge, because the inferences that a prover can draw from a problem description and selected background knowledge provide valuable information. How well these inferences fit the problem description depends strongly on the selected background knowledge. Here it is very important that the selected background knowledge is broad enough but still focused.

6.1. Functional Remote Association Tasks

The benchmark problems we use to evaluate the vector-based selection introduced in this paper are the functional Remote Association Tasks (fRAT) [21] which were developed to measure human creativity. In fRAT, three words like *tulip*, *daisy* and *vase* are given and the task is to find a fourth connecting word, called target word (here *flower*). The words are chosen in such a way that a functional connection must be found between the three words and the target word. To solve these problems, a broad background knowledge is necessary. The solution of the above fRAT task requires the background knowledge that tulips and daisies are flowers and that a vase is a container in which flowers are kept.

The dataset [22] used for this evaluation consists of 48 fRAT tasks. Tab. 1 gives some examples for tasks in the dataset.

6.2. Experimental Results

For an fRAT task consisting of the words w_1, w_2, w_3 and the target word w_t, we first generate a simple goal

$$w_1(w_1) \wedge w_2(w_2) \wedge w_3(w_3) \tag{2}$$

Query Words w_1, w_2 and w_3	Target Word w_t
tulip, daisy, vase	flower
sensitive, sob, weep	cry
algebra, calculus, trigonometry	math
duck, sardine, sinker	swim
finger, glove, palm	hand

Table 1
Examples from the fRAT dataset. Given the three query words, the task is to determine the target word which establishes a functional connection.

Vector-based selection on fRAT		
k	% of tasks with w_t in selection	avg. pos. of target word
5	50%	1.63
10	68.75%	2.70
25	79.17%	4.5
50	87.5%	5.85
100	95.83%	11.15
≥235	100%	17.63

Table 2
Results of selecting with vector-based selection for the 48 fRAT tasks: percentage of tasks where the target word w_t occurs in the axioms selected by vector-based selection. Parameter k corresponds to the number of selected axioms.

using the query words of the tasks as predicate and constant symbols and then select for this goal using different selection strategies. Then we check whether the word w_t occurs in the selected axioms. Since we only want to evaluate selection strategies on commonsense knowledge, we do not use a reasoner in the following experiments and leave that to future work. As background knowledge we use ConceptNet [2] which is a knowledge graph containing broad commonsense knowledge in the form of triples. For this evaluation, we use a first-order logic translation [14] of around 125,000 of the English triples of ConceptNet as knowledge base.

We use both vector-based selection and SInE to select axioms for a goal created for an fRAT task and then check if the target word w_t occurs in the selected axioms. Tab. 2 shows the results for vector-based selection, Tab. 3 for SInE. Note that for vector-based selection the k parameter naturally determines the number of axioms contained in the result of the selection. Since the selected axioms are sorted in descending order with respect to the similarity to the goal in vector-based selection, Tab. 2 furthermore provides the average position of the target word in the selected axioms.

The results for SInE in Tab. 3 show that even for recursion depth 6, were SInE selected 2045.88 axioms on average for an fRAT task, in only 37.5% of the tasks the target word occurred in the selection. Compared to that, the result of vector-based selection of only three axioms already contains the target word in 50% of the tasks. As soon as the vector-based selection selects more than 235 axioms, the target word is contained in the selection for all of the tasks. The Fig. 4 illustrates the relationship between the number of axioms selected and the percentage of target

SInE on fRAT		
rec. depth	% of tasks with w_t in selection	avg. number of selected axioms
1	18.75%	8.92
2	22.92%	51.69
3	29.17%	248.10
4	33.33%	766.52
5	25.42%	1474.00
6	37.5%	2045.88

Table 3
Results of selecting with SInE for the 48 fRAT tasks: percentage of tasks where the target word w_t occurs in the axioms selected by SInE.

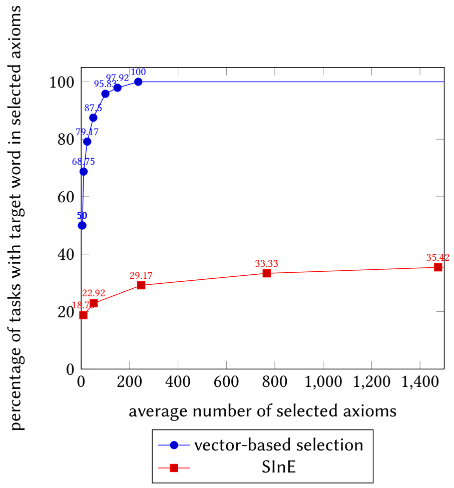

Figure 4: Percentage of fRat tasks for which the target word occurs in the selected axioms depending on the number of selected axioms. Vector-based selection and SInE were used for the selection.

words found for the two selection strategies.

Although SInE selects significantly more axioms than vector-based selection, axioms containing the target word are often not selected. In contrast, vector-based selection is much more focused and even small sets of selected axioms contain axioms mentioning the target word.

The experiments revealed another problem specific for the task of selecting background knowledge from commonsense knowledge bases: Since knowledge bases in this area usually are extremely large, it is reasonable to assume that a user looking for background knowledge for a set of keywords is not aware of the exact symbol names used in the knowledge base. Therefore it can easily happen that a user looks for background knowledge for a set of words which do not coincide with the symbol names used in the knowledge base. For example none of the query words *tulip*, *daisy* and *vase* corresponds to a symbol name in our first-order logic translation of ConceptNet. Therefore a selection using SInE with the goal created from these query words results in an empty selection. In contrast to that, vector-based selection constructs a query vector from the symbol names occurring in the goal (idf-based selection can assume the average idf value for unknown symbols) and selects the k most similar axioms even though the query words from the fRAT task do not occur as symbol names in the knowledge base. As long as the query words occur in the vocabulary of the used word embedding or can be mapped to this vocabulary, it is possible to construct the query vector and select axioms.

The experiments show that vector-based selection is a promising approach for selection on commonsense knowledge. Experiments using reasoners on the selected axioms will be considered in future work.

7. Conclusion and Future Work

Although humans possess large amounts of background knowledge, it is easy for them to focus on the knowledge relevant to a specific problem. Associative reasoning plays an important role in this process. The vector-based selection presented in this paper uses word similarities from word embeddings to model associative reasoning. Our experiments on benchmarks for testing human creativity show that vector-based selection is able to select in a very focused way on commonsense knowledge. In future work, we want to use deductive as well as abductive reasoning on the result of these selections.

In another line of future work, we want to evaluate the usefulness of vector-based selection for the task of solving benchmarks from the commonsense reasoning area like COPA [23].

References

[1] D. Kahneman, Thinking, Fast and Slow, Macmillan, 2011.

[2] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: AAAI, AAAI Press, 2017, pp. 4444–4451.

[3] J. Álvez, P. Lucio, G. Rigau, Adimen-sumo: Reengineering an ontology for first-order reasoning, Int. J. Semantic Web Inf. Syst. 8 (2012) 80–116.

[4] D. B. Lenat, Cyc: A large-scale investment in knowledge infrastructure, Communications of the ACM 38 (1995) 33–38.

[5] K. Hoder, A. Voronkov, Sine qua non for large theory reasoning, in: CADE, volume 6803 of *Lecture Notes in Computer Science*, Springer, 2011, pp. 299–314.

[6] J. Meng, L. C. Paulson, Lightweight relevance filtering for machine-generated resolution problems, J. Applied Logic 7 (2009) 41–57.

[7] A. Roederer, Y. Puzis, G. Sutcliffe, Divvy: An ATP meta-system based on axiom relevance ordering, in: CADE, volume 5663 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 157–162.

[8] G. Sutcliffe, Y. Puzis, SRASS - A semantic relevance axiom selection system, in: CADE, volume 4603 of *Lecture Notes in Computer Science*, Springer, 2007, pp. 295–310.

[9] Q. Liu, Z. Wu, Z. Wang, G. Sutcliffe, Evaluation of axiom selection techniques, in: PAAR+SC^2@IJCAR, volume 2752 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 63–75.

[10] Q. Liu, Y. Xu, Axiom selection over large theory based on new first-order formula metrics, Appl. Intell. 52 (2022) 1793–1807. URL: https://doi.org/10.1007/s10489-021-02469-1. doi:10.1007/s10489-021-02469-1.

[11] S. de Rooij, W. Beek, P. Bloem, F. van Harmelen, S. Schlobach, Are names meaningful? quantifying social meaning on the semantic web, in: ISWC (1), volume 9981 of *Lecture Notes in Computer Science*, 2016, pp. 184–199.

[12] U. Furbach, T. Krämer, C. Schon, Names are not just sound and smoke: Word embeddings for axiom selection, in: CADE, volume 11716 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 250–268.

[13] N. Maslan, M. Roemmele, A. S. Gordon, One hundred challenge problems for logical formalizations of commonsense psychology, in: Twelfth International Symposium on Logical Formalizations of Commonsense Reasoning, Stanford, CA, 2015.

[14] C. Schon, S. Siebert, F. Stolzenburg, Using conceptnet to teach common sense to an automated theorem prover, in: ARCADE@CADE, volume 311 of *EPTCS*, 2019, pp. 19–24.

[15] V. Basile, E. Cabrio, C. Schon, KNEWS: Using Logical and Lexical Semantics to Extract Knowledge from Natural Language, in: Proceedings of the European Conference on Artificial Intelligence (ECAI) 2016 conference, 2016.

[16] J. R. Firth, Papers in Linguistics 1934 - 1951: Rep, Oxford University Press, 1991.

[17] G. A. Miller, W. G. Charles, Contextual correlates of semantic similarity, Language and Cognitive Processes 6 (1991) 1–28. URL: http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ431389.

[18] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: NIPS, 2013, pp. 3111–3119.

[19] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, CoRR abs/1301.3781 (2013). URL: http://arxiv.org/abs/1301.3781. arXiv:1301.3781.

[20] C. Schon, Selection strategies for commonsense knowledge, 2022. URL: https://arxiv.org/abs/2202.09163. doi:10.48550/ARXIV.2202.09163.

[21] P. M. C. Blaine R. Worthen, Toward an improved measure of remote associational ability, Journal of Educational Measurement 8 (1971) 113–123.

[22] A. Olteteanu, M. Schöttner, S. Schuberth, Computationally resurrecting the functional remote associates test using cognitive word associates and principles from a computational solver, Knowl. Based Syst. 168 (2019) 1–9. URL: https://doi.org/10.1016/j.knosys.2018.12.023. doi:10.1016/j.knosys.2018.12.023.

[23] M. Roemmele, C. A. Bejan, A. S. Gordon, Choice of plausible alternatives: An evaluation of commonsense causal reasoning, in: AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning, AAAI, 2011.