Semantic Embeddings for Modal Logic (Extended Abstract)

Pietro Casavecchia[1], Mauro Milella[1], Riccardo Monego[2], Alberto Paparella[1,*], Guido Sciavicco[1] and Leonardo Serrentino[1]

[1]*Department of Mathematics and Computer Science, University of Ferrara, Ferrara, Italy*
[2]*Department of Mathematics, University of Turin, Turin, Italy*

Abstract

The concept of similarity between formulas in formal logic does not have a universally accepted interpretation; intuitively, two formulas should be considered similar if their behaviour is similar, that is, if they agree on most models. In this paper, we propose a novel framework to assess formula similarity inspired by word embeddings in natural language processing. To test the properties of our embedding, we choose modal logic, which provides a natural compromise between propositional and first-order logic. The results of our proof-of-concept implementation suggest that it is possible to represent the behaviour of logical formulas as numerical vectors, regardless of the properties of the underlying frame. Our technique is easily transferable to other logical frameworks, including temporal, spatial, description logics, and, possibly, even first-order frameworks, paving the way for various potential applications.

Keywords

Modal logic, Vector embeddings, Automated reasoning

1. Introduction

In the context of natural language processing, a *word embedding* is a real-valued vector encoding the meaning of a word, so that words close in the vector space are expected to be semantically similar. Analogously, one could consider an embedding of logical formulas, in such a way that the geometric arrangement of vectors reflects the semantics. Such an encoding would also provide a similarity measure for formulas. Similar ideas appeared in [1], introducing an embedding of first-order models (not formulas); in [2], proposing to learn the behaviour of first-order formulas as an embedding; in [3], where first-order Horn clauses are encoded in low-dimension. Conversely, semantic similarity among propositional logic formulas was tackled in [4] and in [5], applied to first-order logic argumentations; [6], learning the representation of predicates using a large language model; and in [7], applied to learning the parameters of signal temporal logic formulas.

In this work, we focus on basic modal logic [8]. Our choice is informed by the following considerations: (i) modal logic is the foundation of many other logical frameworks, also of relevance to applied computer science (concurrency, spatial and temporal reasoning); (ii) basic modal logic (and many of its normal extensions) enjoys the finite model property (namely, every satisfiable formula can be satisfied on a finite model); this is of computational relevance; and (iii) the model checking problem for basic modal logic is polynomial in both the formula length and the model size [9]. Nonetheless, our approach is easily generalizable to other logical frameworks, including many-valued and fuzzy ones.

We propose an algorithm that given $n \in \mathbb{N}$ implements an adversarial search strategy that eventually converges to a family $\{M_1, \ldots, M_n\}$ of n models and a family $\{\varphi_1, \ldots, \varphi_n\}$ of n formulas. This

ICTCS 2026: Italian Conference on Theoretical Computer Science, September 7–9, 2026, Udine, Italy
*Corresponding author.
pietro.casavecchia@unife.it (P. Casavecchia); mauro.milella@unife.it (M. Milella); riccardo.monego@edu.unito.it (R. Monego); alberto.paparella@unife.it (A. Paparella); guido.sciavicco@unife.it (G. Sciavicco); leonardo.serrentino@unife.it (L. Serrentino)
https://alberto-paparella.github.io/ (A. Paparella); https://sites.google.com/unife.it/guido/ (G. Sciavicco)
0009-0000-9117-234X (P. Casavecchia); 0000-0001-7128-6745 (M. Milella); 0009-0008-3936-3753 (R. Monego); 0009-0007-1653-3660 (A. Paparella); 0000-0002-9221-879X (G. Sciavicco); 0009-0004-9884-0194 (L. Serrentino)

mathematical object will be called a *representation base*. Intuitively, the algorithm learns a finite approximation of the space of all formulas and models. Moreover, the representation base induces an embedding of any formula into $\mathbb{R}^n$: for any formula φ, we consider the numeric vector $[M_i(\varphi)]_{i=1}^n$, which in the i-th entry has 1 or 0 depending on whether the model M_i satisfies φ or not. Dually, the final set of formulas induces a similar embedding of a model. With this embedding, the spaces of formulas and models inherit a metric structure, given by their encoding into $\mathbb{R}^n$; the induced similarity measure will be semantic in nature.

Even though this work is intended as a *proof of concept* generalizable to other logics, such a construction has a number of potential applications: (i) semantic evaluation systems for establishing the quality of natural-to-formal language translation approaches; (ii) guiding systems for logic-based machine learning algorithms, useful to avoid exploring too similar areas in the search space; (iii) loss functions for gradient-based symbolic or neuro-symbolic processes; (iv) highly efficient approximation for formal automatic reasoning. Furthermore, for academic goals, basic modal logic is complex enough to provide meaningful results, while still being tractable.

2. Preliminaries

In this section, we express some basic notions regarding modal logic and Kripke semantics. For a more thorough introduction, the reader is referred to [9].

Definition 1. *Let $\mathcal{AP}$ be a vocabulary of* proposition letters. *The well-formed formulas of the* (propositional) modal logic K *are generated by the following grammar:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi, \tag{1}$$

where $p \in \mathcal{AP}$.

The remaining classic Boolean operators and the modal operator $\Box$ can be obtained as shortcuts. The *length* $l(\varphi)$ of a formula φ is the number of its symbols; its *modal depth* $d(\varphi)$ is the maximum number of nested modal operators in the syntax tree of φ; finally, its *height* $h(\varphi)$ is the height of its syntax tree.

Definition 2. *A* Kripke model *over $\mathcal{AP}$ is a triple $M = (W, R, V)$, where W is a set whose elements are called* worlds, *$R \subseteq W \times W$ is a binary relation, called* accessibility relation, *and $V : W \to 2^{\mathcal{AP}}$ is a function called* valuation *which associates to each world w the set of proposition letters $V(w) \subseteq \mathcal{AP}$ that are true at the world w. With abuse, we write $w \in M$ instead of $w \in W$. The* cardinality *$|M|$ of the model is the number of its worlds. For a given model M, a world w and a formula φ, we write $M, w \Vdash \varphi$ to mean that φ is true at the world w in the model M. The truth relation $\Vdash$ is defined inductively on the complexity of formulas:*

$$\begin{array}{lll} M, w \Vdash p & \textit{iff} & p \in V(w), \textit{ for all } p \in \mathcal{AP}; \\ M, w \Vdash \neg\psi & \textit{iff} & M, w \not\Vdash \psi; \\ M, w \Vdash \psi_1 \wedge \psi_2 & \textit{iff} & M, w \Vdash \psi_1 \textit{ and } M, w \Vdash \psi_2; \\ M, w \Vdash \Diamond\psi & \textit{iff} & \textit{there exists } w' \textit{ s.t. } wRw' \textit{ and } M, w' \Vdash \psi. \end{array}$$

We say that a model M satisfies a modal formula φ if and only if $M, w_0 \Vdash \varphi$, where w_0 is a uniquely identified initial *world (this is also denoted by $M(\varphi) = 1$); in this case we say that φ is* satisfiable. *If M does not satisfy φ, we write $M(\varphi) = 0$.*

As with any logical framework, several problems about formulas and their models have been studied. To name a few, beside satisfiability we shall focus on *logical equivalence*, that is, given two formulas φ, ψ, one must decide whether φ and ψ are equivalent, namely, whether φ and ψ are satisfied by exactly the same models (this is denoted by $\varphi \equiv \psi$), and on *logical entailment*, that is, given two formulas φ, ψ, one must decide whether φ entails ψ, namely, whether ψ is satisfied on every model that satisfies φ (this is denoted by $\varphi \Vdash \psi$). In some cases, there is dual version of such problems; for example, given two models

M, N, one wonders whether they satisfy exactly the same formulas, that is, *bisimilar* [9] (denoted $M \cong N$); the dual version of other problems are also possible, but not as common. Interestingly enough, all problems in modal logic share a basic one, known as *model checking*: given a model M and a formula φ, check whether M satisfies φ. Thanks to its applications in formal verification of both hardware and software, the model checking problem has been popularized for more expressive languages, most of which are still rooted in modal logic. In the case of basic modal logic, a classic model checking algorithm can evaluate $M(\varphi)$ in time $O\big(l(\varphi)n^2\big)$, where n is the cardinality of the Kripke model [10].

The basic modal logic K can be extended in several ways, including allowing more than one accessibility relation and diamond operator (*multi-modal* logics), and/or considering fuzzy valuations V and accessibility relations (*many-valued* modal logics), and/or imposing conditions on the accessibility relations, such as transitivity, reflexivity, and so on (these extensions include some of the most popular *temporal* and *spatial* modal logics). We expect that our approach can be extended to all such cases and beyond; here, however, we focus on K only.

3. Semantic E embeddings

The guiding principle of a semantic embedding is that similar formulas should behave similarly on models, but, as we have already noticed, there is a duality between formulas and models. This suggests that a semantic embedding should allow one to treat models and formulas alike. Let $[0;1]$ be the interval of all reals between 0 and 1.

Definition 3. *Given a modal logic $\mathcal{L}$, a* similarity system *is a pair $\mathcal{S} = (\mathcal{S}_f, \mathcal{S}_m)$, where $\mathcal{S}_f$ and $\mathcal{S}_m$ are functions*

$$\mathcal{S}_f\colon \mathbb{L} \times \mathbb{L} \to [0;1], \qquad \mathcal{S}_m\colon \mathbb{M} \times \mathbb{M} \to [0;1]$$

such that they are both reflexive and symmetric, that is

$$\begin{array}{lll} \forall \varphi, \psi \in \mathbb{L}, & \mathcal{S}_f(\varphi,\varphi) = 1 \quad \textit{implies} & \mathcal{S}_f(\varphi,\psi) = \mathcal{S}_f(\psi,\varphi), \\ \forall M, N \in \mathbb{M}, & \mathcal{S}_m(M,M) = 1 \quad \textit{implies} & \mathcal{S}_f(M,N) = \mathcal{S}_f(N,M). \end{array}$$

Among all possible similarity systems, we are interested in *semantic*-based ones.

Definition 4. *Given a modal logic $\mathcal{L}$, a* semantic similarity system *is a similarity system $\mathcal{S} = (\mathcal{S}_f, \mathcal{S}_m)$ in which S_f and S_m are compatible with equivalence and bisimilarity, respectively, that is, whenever φ and ψ are equivalent, $\mathcal{S}_f(\varphi,\psi) = 1$ and whenever M and N are bisimilar, $\mathcal{S}_f(M,N) = 1$.*

Inspired by word embeddings, our approach to semantic similarity for modal logic formulas is based on the idea that formulas and models can be mapped into $\mathbb{R}^n$.

Definition 5. *Given a modal logic $\mathcal{L}$, a* semantic vector embedding of dimension n *is a pair $(\mathcal{T}_f, \mathcal{T}_m)$ of functions*

$$\mathcal{T}_f\colon \mathbb{L} \to \mathbb{R}^n \qquad \mathcal{T}_m\colon \mathbb{M} \to \mathbb{R}^n,$$

such that it preserves equivalence and bisimilarity, that is

$$\begin{array}{llll} \forall \varphi, \psi \in \mathbb{L}, & \textit{if} \quad \varphi \equiv \psi & \textit{then} & \mathcal{T}_f(\varphi) = \mathcal{T}_f(\psi), \\ \forall M, N \in \mathbb{M}, & \textit{if} \quad M \cong N & \textit{then} & \mathcal{T}_m(M) = \mathcal{T}_m(N). \end{array}$$

An obvious way to construct a semantic vector embedding is founded on the following definition.

Definition 6. *Given a modal logic $\mathcal{L}$, a* representation base of dimension n *for $\mathcal{L}$ is a pair $\mathcal{B} = (\mathcal{M}, \Phi)$, where $\mathcal{M} = \{M_1, \ldots, M_n\}$ is a set of n Kripke models and $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ is a set of n formulas. A representation base naturally induces a* basic *vector embedding $(\mathcal{T}_f, \mathcal{T}_m)$ defined by*

$$\mathcal{T}_f(\varphi) = (M_1(\varphi), \ldots, M_n(\varphi)), \qquad \mathcal{T}_m(M) = (M(\varphi_1), \ldots, M(\varphi_n)).$$

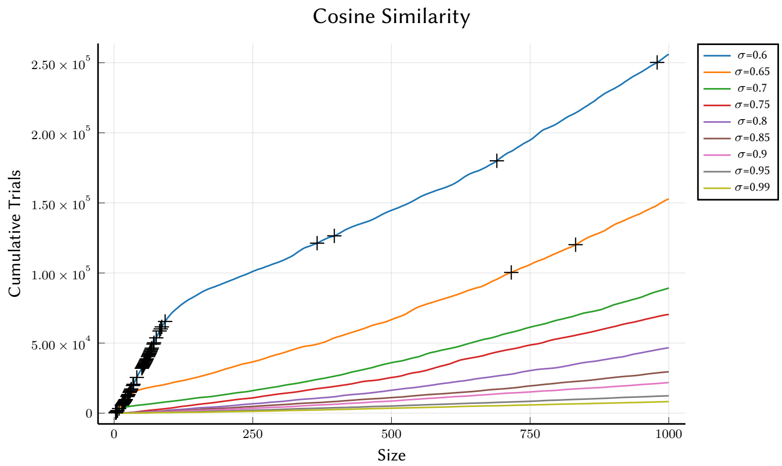

Figure 1: Cumulative sum of all the trials needed to add a new σ-adequate formula/model pair. Crosses highlights runs for which the iteration limit $max = 1000$ is reached.

It is immediate to check that the maps induced by a representation base preserve equivalence and bisimilarity; thus, the vector embedding so obtained is indeed a semantic vector embedding.

A similarity system can be built on a transformation system induced by a representation base. An effective way to obtain it from a basic vector embedding is via a *fuzzy tolerance relation*, that is, a function $F\colon X \times X \to [0;1]$ such that it is reflexive ($F(x,x) = 1$) and symmetric ($F(x,y) = F(y,x)$) [11].

Theorem 1. *Let $\mathcal{L}$ be a modal logic, $\mathcal{B}$ a representation base of size n and $F\colon [0;1]^n \times [0;1]^n \to [0;1]$ a fuzzy tolerance on the unit hypercube of dimension n. Also, let $(\mathcal{T}_f, \mathcal{T}_m)$ be the basic semantic vector embedding induced by $\mathcal{B}$, and define the similarity system $\mathcal{S}$ as*

$$\mathcal{S}_f(\varphi,\psi) = F(\mathcal{T}_f(\varphi), \mathcal{T}_f(\psi)), \qquad \mathcal{S}_m(M,N) = F(\mathcal{T}_m(M), \mathcal{T}_m(N)).$$

Then, $\mathcal{S}$ is a semantic similarity system.

Intuitively, for a basic embedding to be useful, both the models and the formulas should be well spread throughout $\mathbb{M}$ and $\mathbb{L}$ respectively, and not clustered. Unfortunately, there is no formal definition for this notion. One way to express this idea is to require that each pair of vectors in the embedding is *sufficiently dissimilar*. Formally, fix F a fuzzy tolerance relation on the continuous hypercube $[0;1]^n$. For every set of formulas $\Phi = \{\varphi_1, \dots, \varphi_n\}$, we can actively search for a set of models $\mathcal{M} = \{M_1, \dots, M_n\}$ such that, for every pair of vectors in the embedding, $F(\mathcal{T}_m(M_i), \mathcal{T}_m(M_j))$ is lower than some predetermined threshold σ. This set of models $\mathcal{M}$ is not yet guaranteed to properly represent the whole universe $\mathbb{M}$ of all models, since the angles have been measured using only a finite family of formulas. Dually, fixed a set of models $\mathcal{M} = \{M_1, \dots, M_n\}$, we can actively search for a set of formulas $\Phi = \{\varphi_1, \dots, \varphi_n\}$ with the same property, which would represent the same problem, leading us, apparently, to a circular definition. This inspires the following notion, reminiscent of uniformly discrete sets and Delone sets [12].

Definition 7. *Consider a modal logic $\mathcal{L}$, a representation base $(\mathcal{M}, \Phi)$ of size n and a fixed fuzzy tolerance on $[0;1]^n$. Let $(\mathcal{S}_f, \mathcal{S}_m)$ denote the induced semantic similarity system, as in Theorem 1. For $\sigma \in [0;1]$, we say that the representation base $(\mathcal{M}, \Phi)$ is* σ-separated *if and only if, for each pair of distinct formulas and models in the representation base, their similarity is less than σ:*

$$\forall \varphi, \psi \in \Phi, \quad \varphi \neq \psi \quad \textit{implies} \quad \mathcal{S}_f(\varphi,\psi) < \sigma,$$
$$\forall M, N \in \mathcal{M}, \quad M \neq N \quad \textit{implies} \quad \mathcal{S}_m(M,N) < \sigma.$$

4. Experiments

We designed and implemented a prototypical adversarial algorithm [13] to learn a representation base, and then tested the quality of the induced transformation and similarity systems. We fixed the

L, D	Satisfiability			Equivalence			Entailment			Similarity variance		
	$\sigma = 0.6$	$\sigma = 0.65$	$\sigma = 0.7$	$\sigma = 0.6$	$\sigma = 0.65$	$\sigma = 0.7$	$\sigma = 0.6$	$\sigma = 0.65$	$\sigma = 0.7$	$\sigma = 0.6$	$\sigma = 0.65$	$\sigma = 0.7$
4, 1	1	1	0	1	1	1	4	4	3	0.026	0.024	0.002
5, 2	0	0	1	0	0	0	3	3	4	0.028	0.025	0.024
6, 3	2	2	0	1	1	1	4	8	4	0.027	0.026	0.023
7, 4	0	0	0	0	0	0	3	6	4	0.024	0.023	0.020
8, 5	0	0	1	0	0	1	2	2	3	0.024	0.022	0.018
9, 6	2	1	1	0	0	0	7	8	6	0.026	0.024	0.021
10, 7	1	1	0	0	0	0	4	6	5	0.021	0.019	0.017
11, 8	1	1	1	0	0	1	7	7	1	0.023	0.022	0.019
12, 9	1	0	0	1	1	0	6	6	6	0.025	0.022	0.021
13, 10	0	0	0	0	1	0	3	7	7	0.022	0.020	0.017

Table 1
On the left: number of errors out of 500 trials per different problem and vector embedding. On the right: comparison with a randomly generated representation based, specifically, variances out of 500 pairs of formulas.

vocabulary $\mathcal{AP} = \{p, q, r, s, t\}$ and used cosine similarity as fuzzy tolerance.

In Fig. 1, one can see the learning curves of a σ-separate representation base with a maximal modal depth $D = 10$, a maximal overall height $H = 20$, and $\sigma \in [0.6, 1.0]$, step 0.05. During the computation of a base, we check if the σ-separation is broken, and we fix a maximum number max of attempts at fixing the problem; the crosses in Fig. 1 indicate when fixing failed. From the figure, one can see how the number of cumulative attempts grows as we require for more separation. Then, in Tab. 1 (left) for three different embeddings obtained with $\sigma \in \{0.6, 0.65, 0.7\}$, we ask the following questions: given a formula φ (resp., pair of formulas φ, ψ) with a certain length L and modal depth D, how often φ is not satisfied (resp., φ entails ψ, φ and ψ are equivalent) in the corresponding base $\mathcal{B}$ but is satisfiable (resp., does not entail, are not equivalent)? Each line contains number of such misalignments in 500 trials. As it can be seen, such a number is surprisingly low: in the worst case, 8 times out 500, in the transformation base it was the case that a certain φ seemingly entailed a certain ψ and that claim turned out to be false; overall, the total frequency of such errors was slightly over 3%, and much less when considering satisfiability and equivalence only. Finally, in Tab. 1, right, the corresponding three similarity systems are tested for stability: each line contains the variance of the differences between the similarity for two random formulas computed, in each case, in $\mathcal{S}$, and their similarity computed on 1000 random fresh models.

Since our notion of formula similarity is original, the results of our, very initial, experiments may be hard to interpret. The surprisingly low error rate shown by our bases suggests that a similarity system built on such bases should properly represent the general behaviour of of formulas when compared to each other. On the other hand, since there is no ground truth for formula similarity, objectively evaluating this notion is harder, and requires further and more structured experimentation.

5. Conclusions

We proposed a structured concept of semantic similarity between formulas and models, potentially applicable to a wide range of formalism, and initially tested for the modal logic K.

Acknowledgments

This research was partially funded by the INdAM - GNCS Project "Neuro-symbolic semantic contract for rigorous reachability and certificates in cyber-physical systems (CPS)" (CUP: E53C25002010001); Mauro Milella, Alberto Paparella and Guido Sciavicco are members of INdAM.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] T. Sato, Embedding tarskian semantics in vector spaces, in: The Workshops of the The Thirty-First AAAI Conference on Artificial Intelligence, Saturday, February 4-9, 2017, San Francisco, California, USA, volume WS-17 of *AAAI Technical Report*, AAAI Press, 2017. URL: https://aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15072.

[2] T. Rocktäschel, M. Bosnjak, S. Singh, S. Riedel, Low-dimensional embeddings of logic, in: Y. Artzi, T. Kwiatkowski, J. Berant (Eds.), Proceedings of the ACL 2014 Workshop on Semantic Parsing, Association for Computational Linguistics, Baltimore, MD, 2014, pp. 45–49. doi:10.3115/v1/W14-2409.

[3] X. Tang, S.-C. Zhu, Y. Liang, M. Zhang, RulE: Knowledge graph reasoning with rule embedding, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 4316–4335. doi:10.18653/v1/2024.findings-acl.256.

[4] A. Ballout, C. da Costa Pereira, A. G. B. Tettamanzi, Learning to classify logical formulas based on their semantic similarity, in: R. Aydogan, N. Criado, J. Lang, V. Sánchez-Anguix, M. Serramia (Eds.), PRIMA 2022: Principles and Practice of Multi-Agent Systems - 24th International Conference, Valencia, Spain, November 16-18, 2022, Proceedings, Lecture Notes in Computer Science, Springer, 2022, pp. 364–380. URL: https://doi.org/10.1007/978-3-031-21203-1_22. doi:10.1007/978-3-031-21203-1_22.

[5] L. Amgoud, V. David, Measuring similarity between logical arguments, in: M. Thielscher, F. Toni, F. Wolter (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018, AAAI Press, 2018, pp. 98–107. URL: https://aaai.org/papers/13-17999-measuring-similarity-between-logical-arguments/.

[6] N. F. Ariyani, Z. Bouraoui, R. Booth, S. Schockaert, Can language models learn embeddings of propositional logic assertions?, in: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 2766–2776.

[7] G. Saveri, L. Nenzi, L. Bortolussi, J. Kretínský, STL2VEC: Semantic and interpretable vector representation of temporal logic, in: ECAI 2024 - 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain - Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024), Frontiers in Artificial Intelligence and Applications, IOS Press, 2024, pp. 1381–1388. URL: https://doi.org/10.3233/FAIA240638. doi:10.3233/FAIA240638.

[8] J. Garson, Modal Logic, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Spring 2024 ed., Metaphysics Research Lab, Stanford University, 2024.

[9] P. Blackburn, M. d. Rijke, Y. Venema, Modal Logic, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 2001.

[10] E. Clarke, E. Emerson, Design and Synthesis of Synchronization Skeletons Using Branching-Time Temporal Logic, in: Proc. of the 1981 Workshop on Logics of Programs, volume 131 of *LNCS*, 1981, pp. 52–71.

[11] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron, Rough sets: A tutorial, Rough fuzzy hybridization: A new trend in decision-making (1999) 3–98.

[12] A. Abdelkader, D. M. Mount, Economical delone sets for approximating convex bodies, in: 16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2018), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018, pp. 4–1.

[13] S. J. Russell, Artificial intelligence a modern approach, Pearson Education, Inc., 2010.