부정 스키마의 의미론적 양상

태 강 수†

요 약

지능형 에이전트를 구현하는데 있어서 하나의 근본적인 문제는 에이전트가 자신의 인식이나 행동의 의미를 이해하지 못한다는 점이다. 에이전트가 세계를 이해하지 못하는 이유중의 하나는 의미론적 자질을 단순한 문자열로 변환시키는 구문론적 접근방법에서 야기한다. 이 문제를 해결하기 위해 코헨은 에이전트가 자율적으로 자신의 센서와 행동자를 사용하여 환경과 상호작용 함으로써 고급 개념의 기초가 되는 물리적 스키마를 배우는 의미론적 방법을 소개한다. 하지만 코헨은 스키마를 이해하는 것을 가능하게 해주는 상위 계층의 개념소자는 다루지 않는다. 본 논문에서는 부정은 물리적 스키마의 인식을 가능하게 해주는 메타 스키마라는 제안을 하고 부정의 몇 가지 의미론적 양상들을 증명한다.

Semantic Aspects of Negation as Schema

Kang Soo Tae†

ABSTRACT

A fundamental problem in building an intelligent agent is that an agent does not understand the meaning of its perception or its action. One reason that an agent cannot understand the world is partially caused by a syntactic approach that converts a semantic feature into a simple string. To solve this problem, Cohen introduces a semantic approach that an agent autonomously learns a meaningful representation of physical schemas, on which some advanced conceptual structures are built, from physically interacting with environment using its own sensors and effectors. However, Cohen does not deal with a meta level of conceptual primitive that makes recognizing a schema possible. We propose that negation is a meta schema that enables an agent to recognize a physical schema. We prove some semantic aspects of negation.

키워드 : 기계학습(Machine Learning), 계획(Planning), 지식표현(Knowledge Representation), 지식습득(Know Acquisition)

1. Introduction

One of the goals of artificial intelligence is to understand the nature of intelligence and to build systems that exhibit intelligence[4, 13]. A practical problem in attempting to build a truely intelligent system is that an agent does not understand the meaning of its perception or its action[4-6, 16, 17].

In this paper, we first introduce a syntactic approach to knowledge representation that converts semantic features of the objects in the world into syntactic strings. This partially explains why the systems cannot understand the world. Even though we should admit that the syntactic representation of logical inference is very useful when an agent does not understand the world except what is stored in its knowledge base, it is a fundamental weakness of the traditional knowledge representation approach. The issue is how to connect an agent's knowledge representation to meaning rather than to syntactic strings.

To deal with the issue, Cohen suggests a semantical approach that an agent can learn representation directly from interacting with environment using its sensor, motor and language. This approach basically assumes physical schemas as conceptual primitives. Furthermore, he claims that an agent can learn these primitives on its own without human intervention : The agent senses its environment through a collection of sensory streams coming directly from its own sensor rather than through the simulated strings given by a human. Sensation is a meaningful token in a stream. Fluents are states with duration[11]. Cohen claims that fluents are the locus for knowledge, where smallest fluents are just copies of its sensations and complex fulents made of corelated fluents become an abstract entity. A set of fluents can be used to define a class such as a graspable object. While this type of conceptual knowledge

† 정 회 원 : 전주대학교 정보기술컴퓨터공학부 교수
논문접수 : 2001년 11월 5일, 심사완료 : 2001년 12월 18일

is specific to an agent's capabilities, a physical schema is a more general kind of fluent that represents objects or classes across domains. The physical schemas are central to the development of a cognitive agent, forming building blocks for further abstract categories

Finally, we propose that there should be another type of conceptual primitive that enables an agent to recognize a physical schema or a class of objects against the rest of objects in the world. This primitive, called negation, can help an agent implicitly to recognize a partition composed of objects not belonging to the class of interest. We will prove that negation is a meta level of schema and we investigate some aspects of negation.

2. Problems with Syntactic Representation

A knowledge representation language is used to express knowledge in a computer-tractable form. The syntax of the representation language describes how to make well-formed sentences. Simple sentences can be combined by a connective into a complicated sentence. For example, $p \wedge q$ is made by combining conjunctive connective $\wedge$ to simpler sentences p and q, while $\neg p$ is made by combining negative connective $\neg$ to p. In this paper, we are interested in understanding some semantic aspects of the negative connective. We will develop an algorithm to implement this idea on a robotic agent in our next step of research.

The semantics of the language deals with the facts in the world. The sentences refers to the facts. The correspondence between a sentence and a fact is provided by an author's interpretation. Conventionally, the author is limited to a human expert, but it is desirable if the author can be an artificial agent itself. A main goal of Cohen's research is to seek this possibility. The meaning of a sentence is what it describes about the world. While facts are part of the world, the representation of facts is encoded in an agent's mind, even though whether an artificial agent can really possess a mind is a very hard research area related to Strong AI. Note that the agent's reasoning mechanism should operate on the picture or the representation of facts, not on the facts[13, 18].

A semantic causality such that a fact must follow from another fact is reflected by a logical rule that a sentence is *entailed* by another sentence. Logical inference is a process that implements this entailment relation between sentences. Let p refer to a fact F_p and q refer to another fact F_q. If F_q necessarily follows from F_p, then q is entailed from p. This example shows how logic connects a syntactic representation with the world. However, the problem with this type of representation is that the agent still does not understand the world. The correspondence between p and F_p is not learned intrinsically by the agent, but it was created by a human author and is extrinsically given to an artificial agent. Rules and symbols are given as strings and stored in the knowledge base. The agent is only connected to the world in the mind of the creator.

Since a computer system does not know about the world except what appears in the knowledge base, it cannot reason about the world as a human does. To prove a sentence or a goal p with little knowledge on the world, the only possible method that the agent can adopt is to demonstrate that knowledge base entails the sentence p. If p is valid, it does not matter that the agent does not know the world. The conclusion is correct under all the world regardless of the agent's knowledge about the interpretation. While the conclusion is meaningless strings to a system, it is meaningful to a human because he or she knows the interpretation.

However, the advantage of using tautology in logical reasoning comes with its price of the exponential overhead in search. To overcome this overhead, systems should use some type of control rules[1]. Prodigy uses the operator selection or rejection rules[3]. Graphplan, an offspring of Prodigy, uses a control rule called mutex that syntactically detects impossible structures to reduce the search space [2]. We can observe that this syntactic rule reflects the semantic rule that two opposite facts cannot exist in the world simultaneously.

Another tradeoff in using the syntactical approach to knowledge representation is that the system may suffer from a redundancy problem. As mentioned, Graphplan uses mutex to infer inconsistency between two operators or predicates, but the planner does not understand negation, and just treats a negative term as a string of characters. If the negation of a proposition, p, is required, then Graphplan defines a new proposition, say *not-p*, which happens to be equivalent to (*not p*). This kind of seemingly simplistic negative notation may cause a redundancy problem such that a state change is described by two processes, such as *add(not-in(x))* and *del(in(x))*, unless an agent is equipped with a special inference knowledge recognizing

in and *not-in* as opposition. In order to extend Graphplan to handle a negative fact, IPP introduces the negative function *not*. Since *not-p* is not used any more, a negative effect can be uniformly handled as *Add(not p)* rather than as {*Add(not-p)*, *Del(p)*}. Thus, *(not in)* can be used instead of *(not-in)* to negate a fact in a domain[9].

The agent using string representation can also suffers from a noisy problem. Suppose that an agent's arm is empty in the actual world. If the agent's sensors are noisy, the agent may internally believe that its arm is empty and that it is also holding an object at the same time : {*arm -empty*, (*holding x)*}. A machine with incomplete domain knowledge cannot detect that it is an impossible state. Note, however, that a human can infer ~ (*holding x*) at the same time. Even though the process of inferring a negative predicate from a positive predicate seems rather self-obvious to a human, it can be used as crucial control knowledge in a machine[16].

3. Learning Semantic Representation by Interaction

The goal of autonomously learning and understanding representation is extremely hard to achieve. Most AI systems manipulate representations that mean what knowledge engineer intend them to mean. The meanings of representations are exogeneous to the systems. This problem is a fundamental weakness of syntactic approach for representation. Even though the goal of understanding representation is extremely hard to achieve, the best evidence of possibility of building an intelligent entity is a human being. In efforts to build intelligent systems that simulate human cognitive capacities, Searle distinguishes strong AI from weak AI. According to weak AI, the computer gives us a very powerful tool to rigorously formulate and test hypotheses in the study of the intelligence. But strong AI claims that machine can be conscious, and the appropriately programmed computer really is a mind, in the sense that computers can be literally said to understand[14]. Searle denies the possibility of strong AI. However, Cohen claims that meaning can be learned by an agent. If an agent can autonomously learn and understand the meaning of a representation, it will save a lot of work for human. The problems described in the previous section is due to the fact that the agent does not understand the representation of its perception and activities.

Cohen is interested in knowing the origin of conceptual knowledge, the earliest distinctions and classes. Currently most of the intellectual work in AI is done not by programs but by humans, and the work of specifying meaning of a representation is also mostly done by people, not programs. Searle claims that meaning can be learned only by a brain-like machine, and cannot be learned by program. However, Cohen claims that some kinds of meaning can be learned by program. While the origin of concepts is hotly debated, Cohen claims that even though babies' minds have some structure at birth, concepts can be learned without supervision by abstracting over representation of activities. There is converging evidence in psychology, philosophy, linguistics, and robotics that human reasoning relies on a set of conceptual primitives. Mandler claims that image-schematic redescriptions of spatial structure can produce conceptual structures from sensorimotor interactions[10]. The image-schemas are pattern detectors or filters that map sensory streams onto partial representations. These primitives are schematic structures that are grounded in physical interaction with the world. These structures, called physical schemas, describe basic relationships and interactions between an agent and objects, such as moving or pushing. Physical schemas are abstract, domain independent descriptions that are themselves ultimately grounded in the physical processes of moving and applying force.

Ideally, the agent should figure out what a symbol means for itself, not given by us. That way, the robot can learn most of what it knows. Cohen is interested in knowing how an agent can develop its own semantic representation by sensorimotor interaction with the world. Returning to the problem that the agent does not understand its perceptions because they are simulated by strings, Cohen claims that the connection of an agent to the environment should be through its own sensors. Then, the agent's perceptual representation can be grounded on sensors. The agent's direct connection to environment provides representation with meaning. This intrinsic meaning leaned by agent counteracts extrinsic meaning in traditional AI that typically builds systems that do exactly what we want them to do.

Cohen's Baby, a robotic agent embedded in the world, learns representations of objects, activities, and categories using rules similar to the image schemas. When a natural baby is born, it acts and perceives and does little else. The representation, concept, and language must arise from action and perception. In Cohen's approach, an agent assumes

physical schemas as a prior structure and learns the structures based on statistical observation. Baby senses its environment through a collection of streams. Sensation is a meaningful token in a stream. Fluents define objects or activities by abstracting regularities in streams, and fluents become representation stored in memory. Note that while streams are the basis for Baby's sensory experience, fluents are the basis for knowledge. Smallest fluents are just copies of its sensations and they become abstract entity by aggregation.

First, Baby learns scopes, which are pairs of streams that tend to change states simultaneously. By a scope, it knows that the corresponding pair of streams is correlated. Base fluents represent the dependencies that produce the correlation of scopes. Baby learns base fluents that correspond to objects in its environment, such as the red rattle. After Baby learns some base fluents, it starts to form context fluents. The context fluents represent the causal relationships that exist among events that happen one right after the other. A set of fluents can be used to define a class such as a graspable object. This type of conceptual knowledge is relatively specific to an agent's capabilities. A physical schema is a more general kind of fluent that represents objects or classes across domains. This schema are central to the development of a cognitive agent, forming building blocks for further abstract categories and bridging the gap between an agent's sensorimotor behavior and its higher level cognitive skills. As an example of a physical schema, the notion of containment covers many situations at varying level of abstractions, containing a block in a box, containing a sheep within a fence, containing a thought within one's head. The last abstract concept is metaphorically understood by relating it to the process of containing something within a hand[6]. Baby learns a surprising amount given only scopes, base fluents, and context fluents.

4. Negation as Schema

The syntactic aspect of negation is simply to add a negative connective to a predicate or a sentence. Now, we will investigate the semantic aspect of negation in the context of a conceptual primitive. A schema is a conceptual primitive which enables an agent to recognize a pattern or a class of objects. Cohen considered only the physical schema that recognizes a pattern or a class through sensorimotor activities. We will analyze the nature of negation as a type of conceptual primitive which enables an agent to recognize a schema itself and prove some aspects of negation.

The world reveals a great deal of regularity, instead of being a random set of objects. Thus, an ontology, or an organization of objects and actions into categories or classes, is a vital part for a cognitive agent. An object is classified based on its attributes, whether they are objective properties[12] such as *color* and *size* or interactive properties[5] such as *graspable* and *fit-in-my-hand*. An inductive learning program is supposed to learn a function $f(x_i) = y_i$ from data of the form (x_i, y_i) for all i. y_i is called classes and f assigns each x to an appropriate class. When there are only two possible y_i values, the system is called to learn a concept, and each x_i is either a positive or a negative example of the concept. f can be viewed as a definition of the concept[15]. From a semantical view, a concept refers to a set of positive examples that satisfy f. There are lot of machine learning techniques that can learn a concept from a set of data by dividing it into two partitions. Note, however, that their function is mostly limited to learning a concept for the positive data only. We say that the techniques learns a concept on the *level of data*. For example, C4.5 can learn which days are good for playing a game[12].

Here is a question pertaining to our research : What is the concept for the set of negative examples? Currently, no machine learning technique seriously asks this kind of question. For example, C4.5 does not need to learn the concept for days which are not good for playing a game. It is mainly because a machine learning system simply focuses on the efficiency of solving problems and it does not need to concern about probing some relationship existing between the positive data and the negative data. We say that this type of system learns a concept on the meta level. Obviously, the negated concept itself is sometimes rather pretty simple because it refers to the set of objects that do not belong to the class. However, to adjust to the real world, an agent may need much more complex cognitive ability such as recognizing a negated or opposite concept. It is true especially in the area of natural language understanding because a human being tends to compare positive and negative facts together, and furthermore represent a negative fact by a positive predicate rather than by negation[17]. For example, we observe that an elementary stu-

dent should learn by heart the opposite concepts such as *difficult vs easy* or *cold vs hot,* which is closely related with the meaning of *not difficult* or *not cold.*

When an agent recognizes a class of objects as a concept, it differentiate the class from the rest of objects which do not belong to the class. It implies that the agent partitions the world into two classes and it knows to which class an object belongs. If an agent knows the concept for an object, it implies that it also knows that certain object does *not* belong to the concept. Thus, dichotomy is the initial step toward conceptualizing the world.

Suppose the universe U of a domain is partitioned into two sets A and B. The complement of A is the set of elements that belong to the universe U, but do not belong to A. B is the complement of A. Both A and B satisfy the definition of a concept. Thus, the negation relationship is actually the complement relationship because the negation of a concept refers to a set of objects that do not belong to the concept. If the agent knows the concept A, it must know the concept B. The existence of B is necessary for defining or recognizing A in any domain U. B functions as a background. We call B the negation of A. A cognitive agent should possess the mental ability to know the complement relationship. Without this ability, it is impossible to recognize a class. Therefore, the ability to partition the world into two part is a cognitive primitive. Based on this argument, we will prove some aspects of negation.

Theorem 1 : Negation is a schema.

Proof) Suppose there is a cognitive agent. We should prove that negation is a conceptual primitive for the agent. As a base case for induction, suppose there exists only one schema or class in the world for the agent's perception. It implies that there is no ontology for the agent, and it is impossible for the agent to understand or reason about the world. Therefore, there should be more than one class in the world. When there are only two classes for the agent's perception, all the objects are divided into two partitions such that one partition satisfies certain property or function while the other partition does not. Thus, the agent is able to recognize one partition as a concept as long as it also recognize the other partition as the negation for the concept. Since it is impossible to recognize a class without its negation, negation is a conceptual primitive. Thus, it is a schema. □

It is impossible for an agent to reason without its ontology. Having one class, with no organization inside the class, actually means that there is no class to recognize in the world. Relativism is a basic approach in understanding a phenomenon. If we are sitting on a chair, we may not know that we are moving, while a person staying in the space outside the earth can perceive that we are moving along with the earth. Just as we can perceive the fact that we are moving only when we have a view point which is not a part of the earth, we can understand a concept only if there is another concept for comparison and this most primitive concept is the negation of the concept.

Thus, while recognizing a class implies that it should recognize its negation, it will be impossible to recognize any two positive concepts at the same time. To help an agent to recognize a concept, negation suppresses any two concepts within the negative partition from being recognized in an agent's mind at the same time. We prove that negation performs some kind of abstraction. First, let's assume as follows :

Assumption : An agent can recognize only one concept at a time

Theorem 2 : Negation performs a mental abstraction.

Proof) A concept is basically a binary membership function asking whether an object belongs to the positive partition. However, the objects that do not belong to the concept can be actually heterogeneously composed of many different classes within the negative partition. Since an agent cannot recognize two or more concepts at a time, any class of objects that do not belong to the concept cannot be recognized as a concept on its own and it should be ignored. Thus, negation is a mental operation abstracting away the difference among many classes in the negative partition in order to make an agent recognize only one concept. □

If there are 4 colors composed of red, blue, green, and white, in our ontology, and an agent's purpose is to learn the concept of the red color, it is irrelevant to know whether the color of an object is white or blue, as far as it is not red. The difference in the three other colors is suppressed into one concept of not being red. Thus negation is abstracting process for recognition. This capacity of making abstraction seems related with our ability to recognize an opposite relationship[17].

Theorem 3 : Negation is a meta level schema.

Proof) While the schematic structures such as scopes and fluents are grounded on sensorimotor interaction with the world, the negation schema is not directly grounded on physical interaction. Rather, it is a schematic structure on which recognizing a physical schema is based. Since negation does not exist independently but its existence is necessary in recognizing a schema itself, it is a meta schema. □

5. Conclusion

An agent does not understand the meaning of its perception or its action partially because of the syntactical knowledge representation. Cohen claims that a meaningful representation can be learned from physically interacting with environment. Cohen is only interested in learning the physical schemas and does not deal with a meta level of conceptual primitive that makes recognizing a schema possible. We propose that negation is a kind of meta schema. We prove three aspects of negation.

References

[1] Allen, J., Hendler, J., and Tate, A., Readings in Machine Planning, Morgan KaufMann Publishers, 1990.

[2] Brum, A. L. and Furst, M L., Fast Planning through Planning Graph Analysis, Artificial Intelligence 90(1-2) : 281-300, 1997.

[3] Carbonell, J. G., Blythe, J., Etzioni, O., Gil, Y., Knoblock, C., Minton, S., Perez, A., and Wang, X. PRODIGY 4.0 : The Manual and Tutorial, *Technical Report CMU-CS-92-150*, Carnegie Mellon University, Pittsburgh, PA, 1992.

[4] Cohen, P. R., Atkin, M. S., Oates, T., Neo : Learning Conceptual Knowledge by Sensorimotor Interaction with an Environment, *Proceedings of the 6th International Conference on Intelligent Autonomous System*, 2000.

[5] Cohen, P. R. Learning Concepts by Interaction, *Proceedings of the 6th International Conference on Intelligent Autonomous System*, 2000.

[6] Cohen, P. R., Oates, T., Adams, N., Beal, C. R., Robot Baby 2001, *Invited Talk at 12th International Conference on Algorithmic Learning Theory, 2001.*

[7] Fikes, R. and Nilsson, N. STRIPS : A New Approach to the Application of Theorem Proving to Problem Solving, in *Artificial Intelligence*, 2, 1971.

[8] Kaelbling, L. P., Oates, T., Hernandes, N., and Finney, S., Learning in Worlds with Objects, 2001 AAAI Spring Symposium Workshop, Stanford University, 2001.

[9] Koehler, J. Extending Planning Graphs to an ADL Subset, Proc. 4th European Conference on Planning, 1997.

[10] Mandler, J. M., How to build a baby : Conceptual primitives, Psychological Review, 99(4), 1992.

[11] McCarthy, J., Situations, atcions, and causal laws, Stanford Artificial intelligence Project : Memo 2, 1963.

[12] Quinlan, J. R., C4.5 : Programs for Machine Learning, Morgan Kaufmann, 1993.

[13] Russell, S. Norvig, P., Artificial Intelligence : A Modern Approach, Prentice-Hall International, 1995.

[14] Searle, J. R., Minds, Brains, and Programs, Behavioral and Brain Science, 1980.

[15] Shavlik, J. W, and Dietterich, T. G., Readings in Machine Learning, Morgan KaufMann Publishers, 1990.

[16] Tae, K. S., Cook, D. and Holder, L. B. Experimentation-Driven Knowledge Acquisition for Planning, *Computational Intelligence*, 15(3), 1999.

[17] 태강수, "에이전트의 부정에 대한 학습", 정보과학회논문지 : 소프트웨어 및 응용, 27(5), 2000.

[18] Wittgenstein, L., Philosophical Investigations, Macmillan, London, 1953.

태 강 수

email : kstae@jeonju.ac.kr

1983년 전북대학교 영어영문학과(학사)

1991년 University of North Texas 컴퓨터과학과(이학석사)

1991년 미 IBM사 근무

1997년 University of Texas 컴퓨터공학과(공학박사)

1997년~1998년 성덕대학 전자계산학과 전임강사

1998년~현재 전주대학교 정보기술컴퓨터공학부 조교수

관심분야 : 기계학습, 계획, 인공지능, 인지과학