Using CBR for Eliciting Applied Ethics of a Domain

Ali Reza Honarvar [1], Nasser Ghasem-Aghaee [2]

***Abstract*— for evaluating and justifying the behavior of intelligent artificial agents in specific domain from ethical aspect, we should compare agent's behavior to applied ethics of that domain. Existing proposed applied ethics by philosophers are usually abstract and general. We propose a system which can produce non-abstract and concrete codes of ethics, which is appropriate for comparing to agent's behavior in specific domain by using CBR mechanism.**

***Index Terms*— computational ethics, eliciting ethics, artificial ethical agent, case-based reasoning, BDI agent.**

I. Introduction

Robotic system and intelligent artificial agent capabilities have advanced dramatically over the last several decades. These technologies have entered in various filed of human life. We now have artificial intelligent systems and robots that are stronger than humans that are smarter than people in certain cases [1]. The capability of agents to make moral decisions has become important issue, when intelligent agents have developed more autonomous and human-like. The new interdisciplinary research area of "Machine Ethics" is concerned with solving this problem [2,3]. Recently many researchers consider this problem and tried to construct artificial ethical agents in order to increase our trust and confidence in creating autonomous agents acting on our behalf. Our aim in this research is to propose a framework for eliciting applied ethical codes of a domain.

As we use various robotic system and artificial intelligent agent in different fields or domains of human life, we can use applied ethics in that domain for evaluating the functions of agent from ethical perspective. Applied ethics is a term used to describe attempts to use philosophical methods to identify the morally correct course of action in various fields of human life [5].

We proposed a general method for eliciting applied ethics from various fields of human life where artificial intelligent agents have entered. So by using our method it is possible to evaluate the functions of artificial agent or robot by produced codes of applied ethics in that domains.

For eliciting codes of applied ethics in a domain, we use a multi-agent system for simulating the real environment of specified domain. We called this system "Ethics Producer System" (EPA). So the considered domain should have the capacity of simulating by a multi-agent system. In "preliminary and basic concept" section we introduced the basis and preliminary concepts that is necessary for understanding this paper. In section 3 we introduced our framework by an example. The details of our general framework described in section 4 and 5. In section 6 we examine various ethical systems to understand how we can evaluate an action from ethical viewpoint. And finally in section 7 the conclusion and future work are described. In order to understand our method clearly, we first describe our framework by an example then we generalized this example and proposed a general framework.

II. Preliminary and Basic Concept

We need an entity which can situate and act in the context of specific domain In order to elicit applied ethical codes in that domain. This entity should act autonomously, reason proactively and deduce new information about working domain or environment by using information about that domain and the feedbacks of its actions.

Autonomous artificial agent is a computer system that is situated, and that is capable of autonomous action in this environment in order to meet its design objectives. It senses its environment, and acts on it, in pursuit of its own agenda. BDI is a software model developed for programming intelligent agents. At first sight characterized by the implementation of an agent's beliefs, desires and intentions, it actually uses these concepts to solve a particular problem in agent programming.

A multi-agent system is a system composed of multiple interacting intelligent agents. Multi-agent systems can be used to solve problems which are difficult or impossible for an individual to solve [*6*]. Artificial agent has necessary properties for playing the role of introduced entity, which can produce and elicit codes of ethics in specific domain. We use artificial agent and multi-agent system for simulating the entity and the working environment of that entity respectively.

Case based reasoning (CBR) is a well known framework to cope with ill-structured tasks, where no enough domain knowledge is available [*7*]. The main objective of CBR is to address the knowledge acquisition bottleneck. Namely, in CBR the reasoner does not make effort to build an abstract

Manuscript received June 26, 2009.
This work was supported by the Islamic Azad University, Arsanjan Branch (IAUA).

[1] Ali Reza Honarvar is the instructor of Computer Engineering Department, Islamic Azad University, Arsanjan Branch (IAUA), Arsanjan, Fars, Iran.
(E-mail: AliReza_Honarvar@yahoo.co.uk)

[2] Nasser Ghasem-Aghaee is the full professor of Computer Engineering Department, Sheikh Bahaei University and University of Isfahan, Isfahan, Iran.
(E-mail: aghaee@eng.ui.ac.ir)

model for domain knowledge to solve the problem, instead, during the problem solving, it relies on the past similar cases, and attempts to find the appropriate solution for the problem at hand, by modifying the past similar solutions [*8*]. Case-based reasoning is particularly well suited to situations in which domain theory is weak or not easy to formalize [*9*].

Casuist BDI-Agent architecture is a method for constructing artificial ethical agent which extends the power of BDI architecture. The Casuist BDI-Agent architecture combines CBR method in AI and bottom up casuist approach in ethics in order to add capability of ethical reasoning to BDI-Agent [4].

III. ELICITING APPLIED CODES OF SELLING ETHICS

Our aim is to illustrate that simulating a selling/buying system by an Ethics Producer System (EPS) we can elicit the codes of applied ethics in selling domain. In this example we have two kinds of agents: the Seller-Agent (SA) and the Buyer-Agent (BA). The SA is an artificial autonomous agent and the BA is a human which are both situated in EPS. The SA can be a robot in a shop or a software agent in an electronic shop on WWW.

SA has three kinds of information: details of items in the shop, its goal(s), in this example it has one goal: increasing the income, and primary actions which should be perform in order to reach the goal(s). The SA also has three kinds of behavior: negotiation with the BA (B1), using its experiences for acting in similar situations to previous experiences (B2) and reasoning to select appropriate action according to its goals when it does not have appropriate experience (B3).

When any customer (BA) refers to the shop, the SA negotiates with he/she to determine his/her needs. The SA uses a CBR-like method for reasoning. Considering our aim, in this paper we will not discuss the negation step of the SA with the BA and details of CBR phase in more details. So, we used their necessary concepts in abstract. The SA uses the acquired information of BA's needs in negotiation step for retrieving its previous experience in similar situation. If it can find or retrieve any experience, it will use it to act in a similar way if its previous act was ethical. If it cannot find any experience or its previous action is unethical then it switches its behavior to the B3.

According to the B3, if the SA tested an action X in previous situation and it found that this is an unethical action, it will not use it (experience) in similar situation in future. So it will select its choice from other alternative actions depends on the goal(s) and performed negotiations. If we want to map SA's behavior to CBR phase of CBR mechanism we can map the B2 to retrieve and reuse, and the B3 to revise phase of CBR.

Consider a scenario which the SA found that the BA needs the item X with the properties of a, b, c. the SA has item X' with the properties of a, b, d in store. The SA refers to its experiences (Case-Memory), but he will not found any (ethical) experience Similar to this situation. The SA should act according to B3. Considering the goal of the SA, it can probably increase the chance of selling item X' if it conceals concrete details of item X'.

Suppose the SA sold the item X' instead of item X to the BA. After the BA's business with the SA, the BA posts its feedback on its business with the SA to it. The SA should collect necessary information from the BA's feedback to evaluate its action from ethical perspective. We consider various ethical systems in ethics to elicit necessary parameters for ethical evaluation of a situation or action. The discussions of computational models of these parameters are described in section 6. These elements are:

1. The amount of pleasure/displeasure of each affected person by the action x.
2. The duration of pleasure/displeasure of each affected person by the action x.
3. The amount of responsibility of an actor (SA in this example) in the action x.
4. The amount of Voluntariness of an actor on the action x.

To illustrate the effectiveness of these parameters for ethical evaluations of an action we test the first two of these parameters on our example which are the basic core of ethical evaluator of Casuist BDI-Agent architecture [4]. For the sake of simplicity we only considered the affection of the action x on the BA, but for increasing the precision of calculation we should consider every entity that can be affected by the action x.

TABLE 1. POSSIBLE VALUES FOR PARAMETERS 1 AND 2

parameters	Symbolic values	Numerical values
The amount and duration of pleasure/ displeasure	None	0.0
	Very low	0.2
	Low	0.4
	Average	0.6
	High	0.8
	Very high	1.0

From the implementation aspect we should specify how each parameter can be simulated by numerical calculation. Table 1 illustrated each parameter and the possible symbolic and numerical values of them for our example. Remember our mentioned scenario, suppose the BA posted its feedbacks on its business with the SA. These feedbacks can be used for ethical evaluation of a situation. These feedbacks are illustrated in table 2.

TABLE 2. THE FEEDBACKS OF THE BA FOR PARAMETER 1 AND 2

parameters	Symbolic values
The amount of pleasure	Average
The mount of displeasure	High
The interval of pleasure	Average
The interval of displeasure	Low

According to the formula described in section 6 and our assumption that the BA is the only entity that will be affected by SA's action, the result of the calculation is a real number between 0 and 1. The calculated result for this example is 0.04. According to table 3 which describes the meaning of numerical calculation by a meaningful symbol the action of SA in business with the BA is full unethical.

TABLE 3. MEANING OF THE NUMERICAL EVALUATION

Numerical result	Means
[0.0 – 0.2)	Full unethical
[0.2 – 0.4)	Unethical
[0.4 – 0.6)	So So
[0.6 – 0.8)	Ethical
[0.8 – 1.0]	Full ethical

After the ethical evaluation of the action, the SA stores the details of that situation (the negotiation and selected decision) and the calculated evaluations to Case-Memory as a new experience (a case in CBR) for future use. As the experiences of the SA increases incrementally, the Case-Memory fill with various experiences in selling domain which contains codes of applied ethics in that domain. for example according to the mentioned scenario, now the Case-Memory contains one ethical rule in selling ethics which declares " if you want increase your income by selling more items in business, it is full unethical to conceal the concrete information about attributes of an item. "

IV. ETHICS PRODUCER SYSTEM

Ethics Producer System is a multi-agent system which has the capability of producing applied ethics with the aid of Ethics Producer Agent (EPA). This system can be used in any domain that can be simulated by a multi-agent system for producing the applied ethics of that domain. This system has the following parts:

Normal agent: is an artificial/non-artificial agent which has the necessary knowledge of working domain as a triple < B,D,I > , where "B" denotes the beliefs, "D" denotes the desires and "I" denotes the intentions of artificial agent. This artificial agent acts behalf of real agents in specified domain.

Ethics Producer Agent: is an artificial agent who has the duty of producing ethical codes by using CBR-like mechanism

Environment: is a simulation of real environment of specified domain.

Ethical Evaluator: this evaluator taking the current situation of environment, agent's mental attributes (BDI) and feedback of agent's behavior as inputs, then evaluates that status from ethical perspective. This evaluator is based on the evaluation parameters that are described in previous section.

Case-Memory: is a memory that is used for storing ethical codes in the format of cases with three parts. These parts are problem, solution and outcome. Outcome shows the result of applying selected solution in the problem from ethical perspective. The problem describes the situations of environment when solution or decision X is applied by EPA.

Retriever: this part is responsible for retrieving similar cases (experiences) which are similar to the current environment situation and agent's mental attributes. When EPA encounters a problem, it refers to Case-Memory by this component.

After introducing the main structure of Ethics Producer System, in the next section as EPA is the central component of the system for producing ethics, we concentrate on the structure of EPA and its case-base reasoning mechanism for producing applied ethics.

V. ELICITING APPLIED ETHICAL CODES BY EPA

As mentioned before, the main goal of this paper is proposing a general method for producing applied ethics automatically in various fields of human life which can be simulated by a multi-agent system. The process of producing ethical codes is defined in algorithm 1. In this algorithm current beliefs, desires and intentions of EPA at specific moment denoted by a triple < B, D, I > respectively. The current encountered problem and environment status denotes by < E, P >, <B, D, I> and <E, P> which shows the current status of encountered problem are used by EPA to retrieve previous experience in similar situation. If it can find or retrieve any experience, it will use it to act in a similar way if its previous act was ethical. If it cannot find any experience or its previous action is unethical then it should consider its goal(s), <B, D, I> and <E, p>, then selects its choice from other alternative actions. After performing the selected action, the EPA should make an ethical evaluation of its behavior, store its new experience and ethical evaluation of it in the CASE-Memory for future use. As the experiences of the EPA increases incrementally the Case-Memory fills with various experiences in selling domain which contains the codes of ethics in that domain

```
EPA
{
   < B, D, I > = GetCurrentMentalAtributes ( )
          < E, P > = SenseEnvironment ( )
          Case = Retriver ( < E, P > , < B, D, I > )
          If  Case == Null then
            {
                    Operation = Select appropriate action according
                    to <E, B> and <B, D, I>
      }
          Else
                {
                   If retrieved case is ethical then Operation=
                   solution part of retrieved case
                     Else
                          {
                            Operation = Select appropriate action
                            according to <E, B> and <B, D, I>

                          }
                }
          Store problem = "< E, P > , < B, D, I > " , soloution =
          "<Operation>" , result = "<EV>" in Case-Memory
}
```

Algorithm 1. The process of producing ethical codes by the EPA

VI. ETHICAL EVALUATION OF AN ACTION

For ethical evaluation of a situation we considered three novel ethical theories: Hedonistic Act Utilitarianism (HAU), Catholic theory of ethics and Kantian Conception of Duties to self and others.

To our knowledge for the first two of them some computational models proposed. In [10] Gips proposed a simple computational models of HAU where For each person, the algorithm simply computes the product of the pleasure or happiness for each person and a weight assigned each person, and then it adds the individual calculated pleasure to obtain the total pleasure of affected people, In [2] Aderson and Aremn propose a modified version of these equations. For each person, the algorithm simply computes the product of the intensity, the duration, and the probability, to obtain the net pleasure for each person, and then it adds the individual net pleasure to obtain the Total Net Pleasure. In [4] Honarvar and Ghasem-aghaee considered the kind of entity which is affected by the agent's behavior in their computational model of HAU. In this paper we considered some aspects of previous works and used below formulation for calculations in mentioned example which can implement the parameters 1 and 2 for ethical evaluation of an action which is described in section 3. This formulation illustrated as follow:

$$Total\ Net\ Pleasure = \sum(\text{pleasure} \times \text{Duration})\ \text{for each affected individual} - \sum(\text{displeasure} \times \text{Duration})\ \text{for each affected individual}$$

In [11] Hämäläinen described some formulas for calculating the amount of Voluntariness of a person for their action according to Catholic theory of ethics. He considered five factors, which affect on the degree of voluntariness and thus responsibility:

Ignorance: Invincible ignorance, which can be overcome by acquiring the needed knowledge, only lessens the voluntariness. Invincible ignorance destroys it totally.

Passion (strong emotion): antecedent passion, which has spontaneously arisen before the will has acted, lessens freedom, but consequent (deliberately aroused) passion may increase it.

Fear: If fear is motive for acting, it lessens voluntariness.

Violence: The external violence destroys voluntariness, if we withhold consent.

Habit: Responsibility for habitual act depends on, how intentionally it has been acquired, and the amount of effort to get rid of it.

He proposed some Voluntariness rules which affect on the degree of responsibility in her article.

In [12] Diane Jeske proposed three situations where we have ethical duties to our acts. Jeske uses an amalgamation of her view and Kant's view to justify three types of duties: a duty to promote the perfection of others, a duty to promote the happiness of only our own intimates such as our friends and family members, and a duty to promote the happiness of our future selves.

To our knowledge her novel claim which is justified by Kant's view has not formulated by any computational model yet. In future work we will consider Jeske's claim and try to propose a method for determining if an artificial agent has any ethical duty on its actions or not.

VII. CONCLUSION

In this work we proposed a preliminary and imperfect system for producing applied ethics in various field of human life where artificial agents can get control of human's duties. The elicited ethical codes are non-abstract in contrast to abstract ethical codes of applied ethics in any domain that proposed by philosophers. These ethical codes can be used for justifying the behavior of intelligent artificial agent from ethical aspect. The elicited applied ethical codes by EPA in a domain can also be used for constructing an artificial implicit ethical agent which is a machine whose designers have made an effort to design them so that they don't have negative ethical effects, by addressing safety and critical reliability concerns during the design process [13]. The combination of different ethical system for ethical evaluation of a situation can be useful and provides a human-like ethical evaluation of any situation for artificial agent field. In future works we will concentrate on more ethical systems to find appropriate parameters which can be used for ethical evaluation of an artificial agent's actions.

REFERENCES

[1] Arkin, R.C., "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture - Part I: Motivation and Philosophy", Proc. Human-Robot Interaction 2008, Amsterdam, NL, March 2008.

[2] M. Anderson, S. Anderson, and C. Armen, "Toward Machine Ethics: Implementing Two Action-Based Ethical Theories," Proc. AAAI 2005 Fall Symp. Machine Ethics, AAAI Press, pp. 1–16, 2005.

[3] Michael Wooldridge, An introduction to Multiagent Systems, JOHN WILEY & SONS, March 2005.

[4] Ali Reza Honarvar, Naser ghasem Aghaee, " Casuist BDI-Agent: A New Extended BDI Architecture with the Capability of Ethical Reasoning", International Conference on Artificial Intelligence and Computational Intelligence, to be published in LNAI ,Springer Verlag, 2009.

[5] M. Anderson,S. Anderson, "Ethical Healthcare Agents" , Studies in Computational Intelligence ,Springer , vol. 107, pp 233-257, 2008.

[6] Stuart Russell, Peter Norving, Artificial Intelligence A Modern Approach, Prentice Hall Publishing, 2003.

[7] Ian Watson, Applying Case-Based Reasoning, Techniques for Enterprise Systems, Morgan Kaufmann Publishers, 1997.

[8] Pal, S., Shiu, Foundations of Soft Case-Based Reasoning , Wiley-Interscience, 2004.

[9] A. Cordier, B. Fuchs, et A. Mille. Engineering and Learning of Adaptation Knowledge in Case-Based Reasoning. In Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW-06), pages 303–317, 2006.

[10] Gips, Towards the Ethical Robot, Android Epistemology, Cambridge MA: MIT Press, pp. 243—252, 1995.

[11] Wilhelmiina Hämäläinen, "Do computers have conscience? Implementing artificial morality" , available at http://cs.joensuu.fi/~whamalai/articles/ethics.pdf and http://citeseerx.ist.psu.edu .

[12] Diane Jeske, ''Perfection, Happiness, and Duties to Self,'' American Philosophical Quarterly, July 1996.

[13] Wendell Wallach, Colin Allen.: Moral Machines: Teaching Robot Right from Wrong, OXFORD UNIVERSITY press, 2009.