Learning Answer Set Programming Rules for Ethical Machines

Abeer Dyoub[1] Stefania Costantini[1] Francesca A. Lisi[2]

[1]Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila, Italy
Abeer.Dyoub@graduate.univaq.it,Stefania.Costantini@univaq.it

[2]Dipartimento di Informatica &
Centro Interdipartimentale di Logica e Applicazioni (CILA)
Università degli Studi di Bari "Aldo Moro", Italy
FrancescaAlessandra.Lisi@uniba.it

June 21, 2019

Outlines

Ethics in Customer Dealings, Why?

"Now you see, that's EXACTLY what we should try to avoid here."

1. Ethics in customer dealings present the company in a good light and customers will trust.

Ethics in Customer Dealings, Why?

"Now you see, that's EXACTLY what we should try to avoid here."

1. Ethics in customer dealings present the company in a good light and customers will trust.
2. Improve the quality of service and foster positive relationships.

Ethics in Customer Dealings, Why?

"Now you see, that's EXACTLY what we should try to avoid here."

1. Ethics in customer dealings present the company in a good light and customers will trust.
2. Improve the quality of service and foster positive relationships.
3. Many top leading companies have a booklet called code of conduct and ethics and new employees are made to sign it.

Ethics in Customer Dealings,Why?

1. For these reasons companies work on enforcing ethical policies in the work place.

Ethics in Customer Dealings,Why?

1. For these reasons companies work on enforcing ethical policies in the work place.
2. They want to ensure ethical behaviour from their employees especially when dealing with customers.

Ethics in Customer Dealings,Why?

1. For these reasons companies work on enforcing ethical policies in the work place.
2. They want to ensure ethical behaviour from their employees especially when dealing with customers.
3. Monitoring employees ethical behaviour is not an easy task. But is crucial for the company reputation and for gaining customers trust and loyalty.

"And the reason for the return?"

Some examples of ethical codes in customer service

1. Confidentiality
2. Honesty
3. Empathy
4. Non Discrimination
5. Accuracy
6. etc.

The Addressed Problem

To ensure ethical behaviour by employees in the online customer service point. The company want to monitor the dialogue with customers in customer online service chat point to help the managers detect any unethical behaviour from their employees towards customers(with respect to the company's codes of ethics). And then to take counter measures accordingly.

The Problem with codes of ethics implementation(computationaly)

- Codes of ethics in domains such as customer service are mostly abstract general codes, which make them quite difficult to apply. Therefore it is quite difficult if not impossible to define codes in a manner that they maybe applied deductively.

The Problem with codes of ethics implementation(computationaly)

- Codes of ethics in domains such as customer service are mostly abstract general codes, which make them quite difficult to apply. Therefore it is quite difficult if not impossible to define codes in a manner that they maybe applied deductively.
- There are no intermediate rules that elaborate the abstract rules or explain how they apply in concrete circumstances.

The Problem with codes of ethics implementation(computationaly)

- Codes of ethics in domains such as customer service are mostly abstract general codes, which make them quite difficult to apply. Therefore it is quite difficult if not impossible to define codes in a manner that they maybe applied deductively.
- There are no intermediate rules that elaborate the abstract rules or explain how they apply in concrete circumstances.
- It is not possible for experts to define intermediate rules to cover all possible situations to which a particular code applies.

Example

- *Honesty: To ensure the ongoing trust of our customers, we act with honesty. Marketing, advertising and sales activities must describe our offerings and services honestly.*

Example

- *Honesty: To ensure the ongoing trust of our customers, we act with honesty. Marketing, advertising and sales activities must describe our offerings and services honestly.*
- *Accuracy: We shall do all it can to collect, rely and process customer requests and complaints accurately. We shall ensure all correspondence is easy to understand, professional and accurate.*

Example

- *Honesty: To ensure the ongoing trust of our customers, we act with honesty. Marketing, advertising and sales activities must describe our offerings and services honestly.*
- *Accuracy: We shall do all it can to collect, rely and process customer requests and complaints accurately. We shall ensure all correspondence is easy to understand, professional and accurate.*
- Abstract principles such as these seems reasonable and appropriate, but in fact it is very hard to apply them in real-world situations.

Example

- *Honesty: To ensure the ongoing trust of our customers, we act with honesty. Marketing, advertising and sales activities must describe our offerings and services honestly.*
- *Accuracy: We shall do all it can to collect, rely and process customer requests and complaints accurately. We shall ensure all correspondence is easy to understand, professional and accurate.*
- Abstract principles such as these seems reasonable and appropriate, but in fact it is very hard to apply them in real-world situations.
- e.g. how can we precisely define "We shall ensure all correspondence is easy to understand, professional and accurate."?

Proposed Approach

why?

An important question to ask here is how can the company's managers evaluate the ethical behavior of employees in such setting. To achieve this end, and help managers to have detailed rules in place for monitoring the behavior of their employees at customer service for violations of the company's ethical codes,

Proposed Approach

why?

An important question to ask here is how can the company's managers evaluate the ethical behavior of employees in such setting. To achieve this end, and help managers to have detailed rules in place for monitoring the behavior of their employees at customer service for violations of the company's ethical codes,

what?

we propose an approach for generating these detailed rules of evaluation from interactions with customers.

Proposed Approach

why?

An important question to ask here is how can the company's managers evaluate the ethical behavior of employees in such setting. To achieve this end, and help managers to have detailed rules in place for monitoring the behavior of their employees at customer service for violations of the company's ethical codes,

what?

we propose an approach for generating these detailed rules of evaluation from interactions with customers.

How?

Our approach is based on Answer Set Programming (ASP) and Inductive Logic Programming (ILP).

Used Techniques

Before continuing to explain the proposed approach in details, I give a short introduction to the techniques we are using.

Answer Set Programming

- ASP is a logic programming paradigm under answer set (or "stable model") semantics
- ASP features a highly declarative and expressive programming language, oriented towards difficult search problems.
- It has been used in a wide variety of applications in different areas like problem solving, configuration, information integration, security analysis, agent systems, semantic web, and planning.
- In ASP, search problems are reduced to computing answer sets, and an answer set solver (i.e., a program for generating stable models) is used to find solutions.

Answer Set Programming

An answer set Program is a collection of rules of the form,

$$H \leftarrow A_1, \ldots, A_m, not\ A_{m+1}, \ldots, not\ A_n$$

Answer Set Programming

An answer set Program is a collection of rules of the form,

$$H \leftarrow A_1, \ldots, A_m, not\ A_{m+1}, \ldots, not\ A_n$$

Where

- each of A_i's is a literal in the sense of classical logic.
- The left-hand side and right-hand side of rules are called *head* and *body*, respectively.
- A rule with empty body ($n = 0$) is called a *unit rule*, or *fact*.
- A rule with empty head is a *constraint*, and states that literals of the body cannot be simultaneously true in any answer set.
- Unlike other semantics, a program may have several answer sets or may have no answer set,
- each answer set is seen as a solution of given problem,

Inductive Logic Programming

ILP

- ILP is a branch of artificial intelligence (AI) which investigates the inductive construction of logical theories from examples and background knowledge.
- It is the intersection between logic programming and machine learning.

Inductive Logic Programming

ILP

- ILP is a branch of artificial intelligence (AI) which investigates the inductive construction of logical theories from examples and background knowledge.
- It is the intersection between logic programming and machine learning.

General Settings

- Assuming some Examples $E = \{E^+, E^-\}$, background knowledge B.
- Find hypothesis H such that $B \cup H \models E^+$ and $B \cup H \not\models E^-$.
- Mode declarations M restrict Hypothesis space.
- M is either a *modeh(r, s)* or a *modeb(r, s)*,
- s is ground literal, template for literals in head or body of a hypothesis,
- r is an integer, the recall, limits how often the scheme can be used.

Our Proposal

- ASP is used to represent the domain knowledge, the ontology of the domain, and scenarios information.

Our Proposal

- ASP is used to represent the domain knowledge, the ontology of the domain, and scenarios information.
- Rules required for ethical reasoning and evaluation of the agent behavior in a certain scenario are learned using XHAIL, which is a Non-monotonic ILP algorithm.

Our Proposal

- ASP is used to represent the domain knowledge, the ontology of the domain, and scenarios information.
- Rules required for ethical reasoning and evaluation of the agent behavior in a certain scenario are learned using XHAIL, which is a Non-monotonic ILP algorithm.
- The inputs to the system are a series of scenarios(cases) in the form of requests and answers, along with the ethical evaluation of the response considering each particular situation.

Our Proposal - Example

To illustrate our approach, let us consider the following scenario:

a customer contacting the customer service asking for a particular product of the company, and the employee talking about the product characteristics and trying to convince the customer to buy the product. (S)he started saying that the product is environmentally friendly (which is irrelevant in this case), and this is an advantage of their product over the same products of other companies.
The question: is it ethical for the employee to say that?

Our Proposal - Example

To illustrate our approach, let us consider the following scenario:

a customer contacting the customer service asking for a particular product of the company, and the employee talking about the product characteristics and trying to convince the customer to buy the product. (S)he started saying that the product is environmentally friendly (which is irrelevant in this case), and this is an advantage of their product over the same products of other companies.
The question: is it ethical for the employee to say that?

We can form an ILP task for our example:

$$ILP(B, E = \{E^+, E^-\}, M)$$

Background Knowledge B

$$B = \begin{cases} ask(customer, infoabout(productx)). \\ answer(environmentallyFriendly). \\ sensitiveSlogan(environmentallyFriendly). \\ not\_relevant(environmentallyFriendly). \\ answer(xxx). \quad sensitiveSlogan(xxx). \quad not\_relevant(xxx). \\ answer(yyy). \quad sensitiveSlogan(yyy). \quad not\_relevant(yyy). \\ answer(zzz). \quad not\_sensitiveSlogan(zzz). \quad relevant(zzz). \\ answer(eee). \quad not\_sensitiveSlogan(eee). \quad relevant(eee). \\ not\_relevant(X) :- not \quad relevant(X), answer(X). \\ not\_sensitiveSlogan(X) :- not \quad sensitiveSlogan(X), answer(X). \end{cases}$$

Examples: Positive and Negative

E^+

$$E^+ = \begin{cases} example & unethical(environmentallyFriendly). \\ example & unethical(xxx). \\ example & unethical(yyy). \end{cases}$$

E^-

$$E^- = \begin{cases} example & notunethical(zzz). \\ example & notunethical(eee). \end{cases}$$

Mode Declarations M

$$M = \begin{cases} modeh & unethical(+answer). \\ modeb & sensitiveSlogan(+answer). \\ modeb & notsensetiveSlogan(+answer). \\ modeb & notrelevant(+answer). \\ modeb & relevant(+answer). \end{cases}$$

Learning Process

- In the running example, E contains three positive examples and two negative examples which must all be explained.

Learning Process

- In the running example, E contains three positive examples and two negative examples which must all be explained.
- Hypotheses are constructed incrementally by generalising one selected example at a time until all examples are covered.

Learning Process

- In the running example, E contains three positive examples and two negative examples which must all be explained.
- Hypotheses are constructed incrementally by generalising one selected example at a time until all examples are covered.
- For each seed example $e \in E$, this is done by constructing and generalizing a preliminary ground hypothesis K, called a Kernel Set of $B\&e$.

Learning Process

- In the running example, E contains three positive examples and two negative examples which must all be explained.
- Hypotheses are constructed incrementally by generalising one selected example at a time until all examples are covered.
- For each seed example $e \in E$, this is done by constructing and generalizing a preliminary ground hypothesis K, called a Kernel Set of $B\&e$.
- Hypotheses are derived in three steps process:

Step1: Abductive Phase

Head atoms of each Kernel Set are computed

$$\Delta = \begin{cases} unethical(environmentallyFriendly). \\ unethical(xxx). \\ unethical(yyy). \end{cases}$$

Δ is:

- a set of ground facts, where each ground instance is a well typed instance of a clause in the language of M.
- where $B \cup \Delta \models E$

Step2: Deductive Phase

This step computes the body literals of a Kernel Set

$$K = \begin{cases} K1 = unethical(environmentallyFriendly) \leftarrow \\ \qquad sensitiveSlogan(environmentallyFriendly), \\ \qquad not\_relevant(environmentallyFriendly), \\ \qquad answer(environmentallyFriendly). \\ K2 = unethical(xxx) \leftarrow sensitiveSlogan(xxx), not\_relevant(xxx), \\ \qquad answer(xxx). \\ K3 = unethical(yyy) \leftarrow sensitiveSlogan(yyy), not\_relevant(yyy), \\ \qquad answer(yyy). \end{cases}$$

Variabilized Kernel Set K_v

$$K_v = \begin{cases} unethical(V) \leftarrow sensitiveSlogan(V), not\_relevant(V), answer(V). \\ unethical(V) \leftarrow sensitiveSlogan(V), not\_relevant(V), answer(V). \\ unethical(V) \leftarrow sensitiveSlogan(V), not\_relevant(V), answer(V). \end{cases}$$

Step3: Inductive Phase

- computes a compressive theory that subsumes K and entails E w.r.t. B.

Step3: Inductive Phase

- computes a compressive theory that subsumes K and entails E w.r.t. B.
- This is done through actual search for hypothesis which is biased by minimality i.e. preference towards hypothesis with fewer literals.

Step3: Inductive Phase

- computes a compressive theory that subsumes K and entails E w.r.t. B.
- This is done through actual search for hypothesis which is biased by minimality i.e. preference towards hypothesis with fewer literals.
- Thus a hypothesis is constructed by deleting from K_v as many literals (and clauses) as possible while ensuring correct coverage of the examples.

Step3: Inductive Phase

- computes a compressive theory that subsumes K and entails E w.r.t. B.
- This is done through actual search for hypothesis which is biased by minimality i.e. preference towards hypothesis with fewer literals.
- Thus a hypothesis is constructed by deleting from K_v as many literals (and clauses) as possible while ensuring correct coverage of the examples.
- This is done by subjecting K_v to syntactic transformation of its clauses.

Step3: Inductive Phase

- computes a compressive theory that subsumes K and entails E w.r.t. B.
- This is done through actual search for hypothesis which is biased by minimality i.e. preference towards hypothesis with fewer literals.
- Thus a hypothesis is constructed by deleting from K_v as many literals (and clauses) as possible while ensuring correct coverage of the examples.
- This is done by subjecting K_v to syntactic transformation of its clauses.
- The three clauses in K_v produce identical transformations resulting in the same final hypothesis

Final Hypothesis H

$$H = \Big\{ unethical(V) \leftarrow sensitiveSlogan(V), not\_relevant(V), answer(V).$$

Complete Example

Three Cases

Let us now consider our agent having three cases together, the above mentioned case and the following two cases(scenarios) along with a set of examples for each case.

case2

an employee give information about client1 to client2 without checking or being sure that client2 is authorized to be given such information.

case3

a customer contacting customer service asking to buy a certain product x. In this context the customer asks about a similar product of another competitor company which is slightly cheaper. Then the employee, in order to convince the customer to buy their product and not think about the other company product, said that the other company uses substandard materials in their production.

Learned Rules

hypotheses from the above cases

$$H = \begin{cases} unethical(V) \leftarrow sensitiveSlogan(V), \\ \qquad not\_relevant(V), answer(V). \\ unethical(giveinfo(V1, V2)) \leftarrow \\ \qquad context(competitor(V2)), \\ \qquad badinfo(V1), info(V1), company(V2). \\ unethical(tell(V2, infoabout(V2))) \leftarrow \\ \qquad not\_authorized(tell(V1, infoabout(V2))), \\ \qquad client(V1), client(V2). \end{cases}$$

Results

supposing that our agent already have the following rule in its knowledge Base:

$$rule1 = \left\{ unethical(V) \leftarrow not\_correct(V), answer(V). \right.$$

Results

So now our agent has four rules for ethical evaluation (the one that she already has in the background knowledge plus the three learned ones).

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.
- Combining ASP with ILP for modeling ethical agents provides many advantages:

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.
- Combining ASP with ILP for modeling ethical agents provides many advantages:
  - increases the reasoning capability of our agent;

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.
- Combining ASP with ILP for modeling ethical agents provides many advantages:
  - increases the reasoning capability of our agent;
  - promotes the adoption of hybrid strategy that allow both topdown design and bottom up learning via context sensitive adaptation of models of ethical behavior;

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.
- Combining ASP with ILP for modeling ethical agents provides many advantages:
  - increases the reasoning capability of our agent;
  - promotes the adoption of hybrid strategy that allow both topdown design and bottom up learning via context sensitive adaptation of models of ethical behavior;
  - allows the generation of rules with valuable expressive and explanatory power which equips our agent with the capacity to give an ethical evaluation and explain the reasons behind this evaluation.

Conclusion

- we presented an approach that makes use of ASP for ethical knowledge representation and reasoning, and uses inductive logic programming for learning ASP rules needed for ethical reasoning.
- Combining ASP with ILP for modeling ethical agents provides many advantages:
    - increases the reasoning capability of our agent;
    - promotes the adoption of hybrid strategy that allow both topdown design and bottom up learning via context sensitive adaptation of models of ethical behavior;
    - allows the generation of rules with valuable expressive and explanatory power which equips our agent with the capacity to give an ethical evaluation and explain the reasons behind this evaluation.
    - In other words, our method supports transparency and accountability of such models, which facilitates instilling confidence and trust in our agent.

Conclusion

- Furthermore, in our opinion and for the sake of transparency, evaluating the ethical behavior of others should be guided by explicit ethical rules determined by competent judges or ethicists or through consensus of ethicists. Our approach provides support for developing these ethical rules.

Conclusion

- As a matter of fact XHAIL provides an appropriate framework for learning ethical rules for customer service. However XHAIL has some limitations; the problem of scalability.

Conclusion

- As a matter of fact XHAIL provides an appropriate framework for learning ethical rules for customer service. However XHAIL has some limitations; the problem of scalability.
- Furthermore, every time we want to add new cases, XHAIL need to relearn the new hypothesis from the whole set of examples (old ones plus the new added ones).

Conclusion

- As a matter of fact XHAIL provides an appropriate framework for learning ethical rules for customer service. However XHAIL has some limitations; the problem of scalability.
- Furthermore, every time we want to add new cases, XHAIL need to relearn the new hypothesis from the whole set of examples (old ones plus the new added ones).
- Therefore, to cope with large volumes of sequential data and also to cope with ethics change over time, we need an incremental learning technique that is able to revise the old learned hypothesis when a new set of examples arrive.

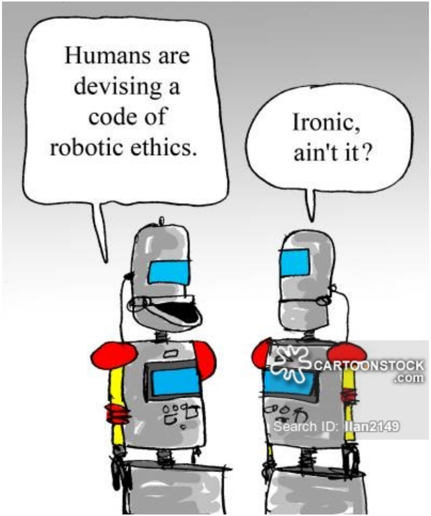
Conclusion

- As a matter of fact XHAIL provides an appropriate framework for learning ethical rules for customer service. However XHAIL has some limitations; the problem of scalability.
- Furthermore, every time we want to add new cases, XHAIL need to relearn the new hypothesis from the whole set of examples (old ones plus the new added ones).
- Therefore, to cope with large volumes of sequential data and also to cope with ethics change over time, we need an incremental learning technique that is able to revise the old learned hypothesis when a new set of examples arrive.
- In fact lately we improved the ethical evaluation capabilities of our agent by using an incremental learning tool(named ILED) to overcome the limitations mentioned above. So our agent can learn incrementally from the interactions with customers to give more accurate evaluations to customer service employees ethical behavior.

Future Work

- We would like to test our agent in a real chat scenario.
- As another future direction we would like to investigate the possibility of judging the ethical behavior from a series of related chat sessions.

The End


Humans are devising a code of robotic ethics.
Ironic, ain't it?
CARTOONSTOCK .com
Search ID: ilan2149