On the question of consistence of the semantic μ-prediction

Stanislav O. Smerdov

Novosibirsk state university
Mechanics and Mathematics Department
The Chair of Discrete Mathematics and Computer Science

Sobolev Institute of Mathematics
Department of Mathematical Logic
Laboratory of Computation Theory and Applied Logic

Adviser: Dr. of Computer Science, Prof. E. E. Vityaev

Sofia 2009

- *The problem of increasing the approximation error while working with uncertaity/fuzziness.*

- *Knowledge of statistical character is captured by probability distributions, truth values are generalized to probabilistic.*

- *One of the major goals of probabilistic (or logical) reasoning consists in explanation/prediction of properties.*

- *The essence of scientific theories lies not so much in their terminology as in general principles of connection between considered sorts of objects.*

- *The problem of increasing the approximation error while working with uncertaity/fuzziness.*
- *Knowledge of statistical character is captured by probability distributions, truth values are generalized to probabilistic.*
- *One of the major goals of probabilistic (or logical) reasoning consists in explanation/prediction of properties.*
- *The essence of scientific theories lies not so much in their terminology as in general principles of connection between considered sorts of objects.*

- *The problem of increasing the approximation error while working with uncertaity/fuzziness.*
- *Knowledge of statistical character is captured by probability distributions, truth values are generalized to probabilistic.*
- *One of the major goals of probabilistic (or logical) reasoning consists in explanation/prediction of properties.*
- *The essence of scientific theories lies not so much in their terminology as in general principles of connection between considered sorts of objects.*

- *The problem of increasing the approximation error while working with uncertaity/fuzziness.*

- *Knowledge of statistical character is captured by probability distributions, truth values are generalized to probabilistic.*

- *One of the major goals of probabilistic (or logical) reasoning consists in explanation/prediction of properties.*

- *The essence of scientific theories lies not so much in their terminology as in general principles of connection between considered sorts of objects.*

There are micro- (logical) and macro- (probabilistic) levels.

Current decisions are made in two-valued classical logic, so consistency of probabilistic theories/predictions (statistical ambiguity problem) is a very important question of AI.

Note that any ϕ should be examined both with its negation: each of them may be specific in prediction of some ψ, e.g. $\mu(\psi \mid \phi) > \mu(\psi \mid \neg\phi)$ or $\mu(\psi \mid \phi) < \mu(\psi \mid \neg\phi)$, where $\mu(\phi) > \mu(\neg\phi)$, for instance.

There are micro- (logical) and macro- (probabilistic) levels.

Current decisions are made in two-valued classical logic, so consistency of probabilistic theories/predictions (statistical ambiguity problem) is a very important question of AI.

Note that any ϕ should be examined both with its negation: each of them may be specific in prediction of some ψ, e.g. $\mu(\psi \mid \phi) > \mu(\psi \mid \neg\phi)$ or $\mu(\psi \mid \phi) < \mu(\psi \mid \neg\phi)$, where $\mu(\phi) > \mu(\neg\phi)$, for instance.

There are micro- (logical) and macro- (probabilistic) levels.

Current decisions are made in two-valued classical logic, so consistency of probabilistic theories/predictions (statistical ambiguity problem) is a very important question of AI.

Note that any ϕ should be examined both with its negation: each of them may be specific in prediction of some ψ, e.g.
$\mu(\psi \mid \phi) > \mu(\psi \mid \neg\phi)$ or $\mu(\psi \mid \phi) < \mu(\psi \mid \neg\phi)$,
where $\mu(\phi) > \mu(\neg\phi)$, for instance.

Language

Let $\mathfrak{L}$ be a first-order language of a finite signature.

Allow literals (not only atoms) to appear in a classical logic programming structures of rule, fact and query; denote corresponding sets as $\mathtt{Rule}_{\mathfrak{L}}$, $\mathtt{Fact}_{\mathfrak{L}}$ and $\mathtt{Query}_{\mathfrak{L}}$.

definition

Binary relation $\mathrm{C}_1 \succ \mathrm{C}_2$ (read "C_1 *is more general than* C_2") between $\mathrm{C}_1 \equiv (\mathrm{A}_1 \Leftarrow \mathrm{B}_1, ..., \mathrm{B}_n), \mathrm{C}_2 \equiv (\mathrm{A}_2 \Leftarrow \mathrm{D}_1, ..., \mathrm{D}_m)$ in $\mathtt{Rule}_{\mathfrak{L}}$ takes place iff there exist a substitution θ such that $\{\mathrm{B}_1\theta, ..., \mathrm{B}_n\theta\} \subseteq \{\mathrm{D}_1, ..., \mathrm{D}_m\}$, $\mathrm{A}_1\theta \equiv \mathrm{A}_2$ and $\not\vdash \mathrm{C}_1 \equiv \mathrm{C}_2$.

Probability over ground sentences

- Let $\mathfrak{G}^*$ be a countable class of observed first-order structures appearing in practice; $\mathfrak{G} \subset \mathfrak{G}^*$ is *a general sampling* consisting of well-studied models.
- Being given $\mathfrak{G}$ we compute a probability measure P over $\mathfrak{G}^*$ with some trusting interval value $\varepsilon > 0$ (according to Kolmogorov); here mathematical statistics is applied.
- Assume $\mu(\phi) \leftrightharpoons \mathrm{P}(\{\mathfrak{A} | \mathfrak{A} \models \phi\})$, where ϕ is a closed formulae.

Probability over ground sentences

- Let $\mathfrak{G}^*$ be a countable class of observed first-order structures appearing in practice; $\mathfrak{G} \subset \mathfrak{G}^*$ is *a general sampling* consisting of well-studied models.

- Being given $\mathfrak{G}$ we compute a probability measure P over $\mathfrak{G}^*$ with some trusting interval value $\varepsilon > 0$ (according to Kolmogorov); here mathematical statistics is applied.

- Assume $\mu(\phi) \leftrightharpoons \mathrm{P}(\{\mathfrak{A} | \mathfrak{A} \models \phi\})$, where ϕ is a closed formulae.

Probability over ground sentences

- Let $\mathfrak{G}^*$ be a countable class of observed first-order structures appearing in practice; $\mathfrak{G} \subset \mathfrak{G}^*$ is *a general sampling* consisting of well-studied models.

- Being given $\mathfrak{G}$ we compute a probability measure P over $\mathfrak{G}^*$ with some trusting interval value $\varepsilon > 0$ (according to Kolmogorov); here mathematical statistics is applied.

- Assume $\mu(\phi) \risingdotseq \mathsf{P}(\{\mathfrak{A} | \mathfrak{A} \vDash \phi\})$, where ϕ is a closed formulae.

Probability over rules

Let $\Theta^{\rm o}$ be a set of all ground substitutions.

Probability of a ground instance of rule is defined as conditional
$$\mu\left(\mathrm{A} \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n\right) = \mu\left(\mathrm{A} \,|\, \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n\right) = \frac{\mu(\mathrm{A} \wedge \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n)}{\mu(\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n)}$$

$$\mathtt{Rule}^{\mu}_{\mathfrak{L}} \rightleftharpoons \{\mathrm{C} \,|\, \textit{for some } \theta \in \Theta^{\rm o} \textit{ probability of } \mathrm{C}\theta \textit{ is determined}\}$$

$$\underline{\mu}\left(\mathrm{C}\right) \rightleftharpoons \inf\left\{\mu\left(\mathrm{C}\theta\right) \,|\, \theta \in \Theta^{\rm o} \textit{ and } \mathrm{C}\theta \in \mathtt{Rule}^{\mu}_{\mathrm{L}}\right\},$$
where $\mathrm{C} \in \mathtt{Rule}^{\mu}_{\mathfrak{L}}$

Best prediction rule

$\mathtt{Fact}_{\mathrm{o}}$ is a set of ground atoms allowing verification in any $\mathfrak{B} \in \mathfrak{G}^*$; *a complete set of alternatives* is

$$\mathtt{Fact}_{\mathrm{o}}^* = \mathtt{Fact}_{\mathrm{o}} \cup \{\neg \mathrm{A} \,|\, \mathrm{A} \in \mathtt{Fact}_{\mathrm{o}}\}$$

definition (E.E. Vityaev, S.O. Smerdov)

A rule $\mathrm{C} \equiv (\mathrm{A} \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n)$ is called *the best prediction rule for some literal* H iff the following conditions are hold:

$i.$ there exist $\theta \in \Theta^{\mathrm{o}}$ such that $\mathrm{A}\theta \equiv \mathrm{H}\theta$, $\{\mathrm{B}_1\theta, \ldots, \mathrm{B}_n\theta\} \subseteq \mathtt{Fact}_{\mathrm{o}}^*$, $\mu\left((\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n)\,\theta\right) \neq 0$ and $\underline{\mu}\,(\mathrm{C}) > \mu\,(\mathrm{H}\theta)$;

$ii.$ maximum of $\underline{\mu}\,(\cdot)$ is achieved on C among rules satisfying (i);

$iii.$ it can't be generalized without loosing $(i\text{–}ii)$.

- Best rules can be viewed as a result of so called *semantic μ-prediction* (notion was introduced in works of E.E. Vityaev) of different literals. For each best rule C used in prediction of some ground H we consider all $\mathrm{C}\theta$ such that θ is a ground substitution satisfying the point (i) of definition.

- We denote by $\mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ the obtained set of described ground instances (over all literals H).

- $\mathtt{Data}\,(\mathfrak{B})$ is a set of actual facts for 1-st order model $\mathfrak{B} \in \mathfrak{G}^{*}$, i.e. consistent subset of $\mathtt{Fact}_{\text{o}}^{*}$ (not necessary maximal).

definition

A set of literals S is called *μ-concurred* iff $\mathsf{P}\,(\{\mathfrak{A}\,|\,\mathfrak{A} \models S\}) \neq 0$.

- Best rules can be viewed as a result of so called *semantic μ-prediction* (notion was introduced in works of E.E. Vityaev) of different literals. For each best rule C used in prediction of some ground H we consider all $\mathrm{C}\theta$ such that θ is a ground substitution satisfying the point (i) of definition.

- We denote by $\mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ the obtained set of described ground instances (over all literals H).

- $\mathtt{Data}\,(\mathfrak{B})$ is a set of actual facts for 1-st order model $\mathfrak{B} \in \mathfrak{G}^*$, i.e. consistent subset of $\mathtt{Fact}_{\mathrm{o}}^*$ (not necessary maximal).

definition

A set of literals S is called *μ-concurred* iff $\mathrm{P}\left(\{\mathfrak{A} \,|\, \mathfrak{A} \models S\}\right) \neq 0$.

- Best rules can be viewed as a result of so called *semantic μ-prediction* (notion was introduced in works of E.E. Vityaev) of different literals. For each best rule C used in prediction of some ground H we consider all $\mathrm{C}\theta$ such that θ is a ground substitution satisfying the point (i) of definition.

- We denote by $\mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ the obtained set of described ground instances (over all literals H).

- $\mathtt{Data}\,(\mathfrak{B})$ is a set of actual facts for 1-st order model $\mathfrak{B} \in \mathfrak{G}^*$, i.e. consistent subset of $\mathtt{Fact}_{\mathrm{o}}^{*}$ (not necessary maximal).

definition

A set of literals S is called *μ-concurred* iff $\mathrm{P}\left(\{\mathfrak{A} \,|\, \mathfrak{A} \vDash S\}\right) \neq 0$.

• Best rules can be viewed as a result of so called *semantic μ-prediction* (notion was introduced in works of E.E. Vityaev) of different literals. For each best rule C used in prediction of some ground H we consider all $\mathrm{C}\theta$ such that θ is a ground substitution satisfying the point (i) of definition.

• We denote by $\mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ the obtained set of described ground instances (over all literals H).

• $\mathtt{Data}\,(\mathfrak{B})$ is a set of actual facts for 1-st order model $\mathfrak{B} \in \mathfrak{G}^*$, i.e. consistent subset of $\mathtt{Fact}_{\mathrm{o}}^*$ (not necessary maximal).

definition

A set of literals S is called *μ-concurred* iff $\mathsf{P}\left(\{\mathfrak{A} \,|\, \mathfrak{A} \vDash S\}\right) \neq 0$.

theorem

Let some ground atom H *be semantically* μ*-predicted by ground instance* $\mathrm{C}_{pos} \in \mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ *of the best rule* C_1 $(\mathrm{C}_{pos} \equiv \mathrm{C}_1\theta_{pos})$*, while* $\neg\mathrm{H}$ *is predicted by* $\mathrm{C}_{neg} \in \mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ $(\mathrm{C}_{neg} \equiv \mathrm{C}_2\theta_{neg})$*. Then the set of atoms from premises of* C_{pos} *and* C_{neg} *is not* μ*-concurred.*

Denote by $\Gamma_{\mathfrak{B}}$ the following set of rules and data

$$\left\{\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \rightarrow \mathrm{A} \,|\, \mathrm{A} \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \in \mathtt{Prdct}_{\mathrm{L}}^{\mu,0}\right\} \cup \mathtt{Data}\,(\mathfrak{B})$$

theorem

Let $\mathtt{Data}\,(\mathfrak{B})$ *be* μ*-concurred. Then minimal theory containing* $\Gamma_{\mathfrak{B}}$ *is logically consistent.*

theorem

Let some ground atom H *be semantically* μ*-predicted by ground instance* $\mathrm{C}_{pos} \in \mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ *of the best rule* C_1 $(\mathrm{C}_{pos} \equiv \mathrm{C}_1\theta_{pos})$*, while* $\neg\mathrm{H}$ *is predicted by* $\mathrm{C}_{neg} \in \mathtt{Prdct}_{\mathfrak{L}}^{\mu,0}$ $(\mathrm{C}_{neg} \equiv \mathrm{C}_2\theta_{neg})$*. Then the set of atoms from premises of* C_{pos} *and* C_{neg} *is not* μ*-concurred.*

Denote by $\Gamma_{\mathfrak{B}}$ the following set of rules and data

$$\left\{\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \rightarrow \mathrm{A} \,\middle|\, \mathrm{A} \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \in \mathtt{Prdct}_{\mathrm{L}}^{\mu,0}\right\} \cup \mathtt{Data}\,(\mathfrak{B})$$

theorem

Let $\mathtt{Data}\,(\mathfrak{B})$ *be* μ*-concurred. Then minimal theory containing* $\Gamma_{\mathfrak{B}}$ *is logically consistent.*

theorem

Let some ground atom H *be semantically* μ*-predicted by ground instance* $\mathrm{C}_{pos} \in \mathtt{Prdct}^{\mu,0}_{\mathfrak{L}}$ *of the best rule* C_1 $(\mathrm{C}_{pos} \equiv \mathrm{C}_1\theta_{pos})$*, while* $\neg\mathrm{H}$ *is predicted by* $\mathrm{C}_{neg} \in \mathtt{Prdct}^{\mu,0}_{\mathfrak{L}}$ $(\mathrm{C}_{neg} \equiv \mathrm{C}_2\theta_{neg})$*. Then the set of atoms from premises of* C_{pos} *and* C_{neg} *is not* μ*-concurred.*

Denote by $\Gamma_{\mathfrak{B}}$ the following set of rules and data

$$\left\{\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \to \mathrm{A} \,\middle|\, \mathrm{A} \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n \in \mathtt{Prdct}^{\mu,0}_{\mathrm{L}}\right\} \cup \mathtt{Data}\,(\mathfrak{B})$$

theorem

Let $\mathtt{Data}\,(\mathfrak{B})$ *be* μ*-concurred. Then minimal theory containing* $\Gamma_{\mathfrak{B}}$ *is logically consistent.*

Thank you for attention.