Game Semantical Rules for Vague Proportional Quantifiers

Matthias Hofer

Supported by FWF projects I1897-N25 and W1255-N23

May 22, 2016

Natural Language Understanding - Is that a hard problem?

or: can we solve it straithforwardly?

three different attempts:

- babelfish for computers
- statistical approaches
- classical logic

and why they do not work:

- sadly, does not exist
- no core semantics
- too rigid

Our fraction of the problem
syntax and examples

Natural language statements with vagueness:

quantifier(range,scope) $\longrightarrow$ **Q(R,S)**

"Almost all students passed the exam."

"About half young mothers live in very small apartments."

"Many rich people live near the city center."

$\mathcal{G}$iles's game for Łukasiewicz logic

our game semantics

- two player (**P**,**O**) zero sum game of perfect information
- states: multisets of formulas $[F_1, \ldots, F_n \mid G_1, \ldots, G_m]$
- states decompose into final states: $[A_1, \ldots, A_{n'} \mid B_1, \ldots, B_{m'}]$
- risk at final state:

  $\langle A_1, \ldots, A_{n'} \mid B_1, \ldots, B_{m'} \rangle = \sum_{1 \leq i \leq m'} \langle B_i \rangle - \sum_{1 \leq j \leq n'} \langle A_j \rangle$

example:

Game Rule ($R_{\rightarrow}$)

If **P** *asserts $F \rightarrow G$ then* **O** *may attack by asserting F, obliging* **P** *to assert G.*

$$[\mid F \rightarrow Q] \longrightarrow [F \mid Q]$$

$$\langle F \mid Q \rangle = \langle Q \rangle - \langle F \rangle = (1 - v_I(Q)) - (1 - v_I(F))$$

$$v_I(F \rightarrow Q) = min(0, 1 - v_I(F) + v_I(Q))$$

Strong Conjunction and Limited Liability

and the adequateness theorem for Ł logic

Game Rule ($R_\&$)

If **P** *asserts* $F\&G$ *then, if* **O** *attacks,* **P** *has to either assert* F *as well as* G*, or else* $\bot$.

Theorem

For every atomic formula A *let* $\langle A \rangle$ *be its risk and let* I *be the Ł-interpretation given by* $v_I(A) = 1 - \langle A \rangle$. *Then, if both,* **P** *and* **O***, play rationally, any game starting in state* $[|F]$ *will end in a state where* **P***'s final risk is* $1 - v_I(F)$.

semi-fuzzy/fully-fuzzy and unary/binary

or: base case restrictions

- semi fuzzy: **quantifier** is fuzzy, but its **arguments** classical
- fully fuzzy: **quantifier** as well as its **arguments** are fuzzy

Game Rule ($R_\forall$)

If **P** *asserts* $\forall x \hat{F}(x)$ *then* **O** *may attack by choosing some constant* c *and* **P** *has to reply by asserting* $\hat{F}(c)$.

Game Rule ($R^2_\forall$)

If **P** *asserts* $all_x(\hat{F}(x), \hat{G}(x))$ *then* **O** *may attack by picking some constant* c *and asserting* $\hat{F}(c)$, *thereby forcing* **P** *to assert* $\hat{G}(c)$.

Proportion and Tolerance

extensional vagueness

Proportion of A's that are B's:

"Almost all planets are uninhabited."

"At most about a third sand corns are black."

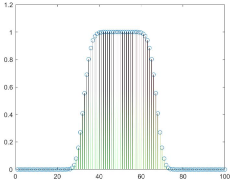
"About half of all humans are women."

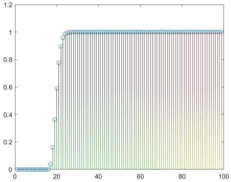
quantifier + tolerance → vague quantifier

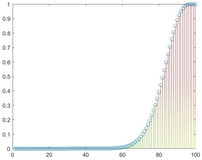
(Vague) Proportional Quantifiers

pictures show truth functions

- “all” / “exists”
- “almost all” / “nearly none”
- “about half” / “about $x\%$”
- “at least about $x\%$” / “at most about $x\%$”







Randomizing Game Rules

one old and one new

Game Rule ($R_{\sqcap}$)

If **P** *asserts* $\sqcap x F(x)$ *then* **P** *has to assert* $F(c)$ *for a randomly picked* c.

$$v_I(\sqcap x \hat{F}(x)) = \frac{\sum_{c \in U} v_I \hat{F}(c)}{|U|} = Prop_x \hat{F}(x)$$

Game Rule ($R_{\sqcap^{j,k}}$)

If **P** *asserts* $\sqcap^{j,k} x \hat{F}(x)$ *then, if* **O** *attacks,* **P** *has to assert* $\hat{F}(c)$ *for* j *of* k *different and uniformly chosen* c*'s, or else* $\bot$.

$$v_I(\sqcap^{j,k} x \hat{F}(x)) = \frac{\sum_{i=j}^{k} \binom{np}{i}\binom{n-np}{k-i}}{\binom{n}{k}}$$

“almost all” and t-process

how we define quantifiers

Game Rule (“almost all”)

If **P** *asserts almost all* $x\hat{F}(x)$*, then, if* **O** *attacks,* t *gets chosen through applying a t-process, and then* $(t+j)$ **different** *constants get picked randomly, and then* **P** *has to assert* $\hat{F}(c)$ *for* j *of them.*

t-process for domain size n:

- $\mathbb{P}^{(n)}(X_t = i) > \mathbb{P}^{(n)}(X_t = i+1)$ for $i \in \{1, ...n-2\}$, $n \geq 3$
- $\mathbb{P}^{(n)}(X_t = i) > \mathbb{P}^{(n+1)}(X_t = i)$ for $i \in \{1, ...n-1\}$, $n \geq 2$
- $\sum_{i=1}^{r-1} i\mathbb{P}^{(n)}(X_t = i) < \sum_{i=1}^{r} i\mathbb{P}^{(n+1)}(X_t = i)$

More Quantifiers Defined

one way to do it

- $Q_j^{[\gtrapprox q]} = \sqcap^{j, \lfloor n(1-q) \rfloor + j}, \qquad j \in \{1, ..., \lfloor nq \rfloor\}$ and $q \in (0,1)$
- $Q_j^{[\lessapprox q]} = \sqcap^{j, \lfloor nq \rfloor + j} \neg, \qquad j \in \{1, ..., \lfloor n(1-q) \rfloor\}$ and $q \in (0,1)$
- $Q_j^{[\backsimeq \frac{1}{2}]} = Q_j^{[\gtrapprox \frac{\lfloor \frac{n}{2} \rfloor}{n}]} \; \& \; Q_j^{[\lessapprox \frac{\lceil \frac{n}{2} \rceil}{n}]}, \qquad j \in \{1, ..., \lfloor \frac{n}{2} \rfloor\}$
- $Q_j^{[\backsimeq q]} = Q_j^{[\gtrapprox q]} \; \& \; Q_j^{[\lessapprox q]}, \; j \in \{1, ..., \min(\lfloor qn \rfloor, \lfloor n(1-q) \rfloor) - 1\}$
- $Q_{t,j}^{[\approx q]} = Q_j^{[\gtrapprox \frac{\lfloor qn \rfloor - t}{n}]} \; \& \; Q_j^{[\lessapprox \frac{\lfloor qn \rfloor + t}{n}]}$, with

tolerance $t \in \{1, ..., \min(\lfloor qn \rfloor - 1, \lfloor n(1-q) \rfloor - 1)\}$,
and rigor $j \in \{1, ..., \min(\lfloor qn \rfloor - t, \lfloor n(1-q) \rfloor - t - 1\}$

Adequateness theorem for $\Pi^{j,k}$

sketch of proof

Theorem

A $Ł(R_{\Pi^{j,k}})$-sentence $\Pi^{j,k} x \hat{F}(x)$, for a classical formula $\hat{F}$ is evaluated to $v_I(\Pi^{j,k} x \hat{F}(x)) = x$ in an interpretation I iff every $\mathcal{G}$-game for $\Pi^{j,k} x \hat{F}(x)$ augmented by rule $(R_{\Pi^{j,k}})$ is $(1-x)$-valued for me under risk value assignment $\langle \cdot \rangle_I$

Proof

For $\langle \Gamma \mid \Delta, \Pi^{j,k} x \hat{F}(x) \rangle = \langle \Gamma \mid \Delta \rangle + \langle \mid \Pi^{j,k} x \hat{F}(x) \rangle$, we only show:

$$\langle \Pi^{j,k} \hat{F}(x) \rangle_I = 1 - v_I(\Pi^{j,k} x \hat{F}(x)) = \frac{\sum_{i=0}^{j-1} \binom{np}{i}\binom{n-np}{k-i}}{\binom{n}{k}}$$

For this, note that:

$$\frac{\sum_{i=0}^{k} \binom{np}{i}\binom{n-np}{k-i}}{\binom{n}{k}} = 1$$

Game Rules for Intensional Quantifiers

many and few

Sets of comparison classes:

$S_{\text{int}} := \{ic_1, ic_2, .., ic_n\}$, and $S^{\text{sa}}_{\text{ext}} := \{ec_1, ec_2, .., ec_n\}$

Game Rule ($R^{\text{ext}}_{\text{Many}}$)

If **P** *asserts* $Q^{many}x\hat{F}(x)$ *then, if* **O** *attacks, some element* $ec_i \in S^{\text{sa}}_{\text{ext}}$ *is chosen randomly, and if* $p^{\text{sa}}_{\hat{F}} \geq p^{\text{sa}}_{i}$ *is true,* **P** *has to replace* $Q^{many}x\hat{F}(x)$ *with one distinguished true atom in his/her tenet, and otherwise with a wrong one.*

Game Rule ($R^{\text{int}}_{\text{Many}}$)

If **P** *asserts* $Q^{many}x\hat{F}(x)$ *then, if* **O** *attacks, some element* $ic_j \in S_{\text{int}}$ *is chosen randomly, and if* $p^{\text{sa}}_{\hat{F}} > p^{j}_{\hat{F}}$ *is true,* **P** *has to replace* $Q^{many}x\hat{F}(x)$ *with one distinguished true atom in his/her tenet, and otherwise with a wrong one.*

How can we apply this approach to real world challenges?

querying and data summarization

Consider:

binary vague quantifier, unary range predicate, any scope formula:

$$Qx(P(x), F(x, \overrightarrow{y}))$$

Example:

$$[\gtrapprox \tfrac{2}{5}]x(\mathrm{LEGO}(x), \mathrm{produced\_in}(x, y_1) \wedge \mathrm{sold\_in}(x, y_2))$$

Natural Language Query:

"Where are at least about 40% of all LEGOs produced and sold?"

conclusion and outlook

an agenda

summary

- natural language understanding and game semantics
- vague proportional quantifiers
- intensional quantifiers

challenges

- fully fuzzy
- imperfect information
- time dependence
- querying and data summarization

the end

for now...

Thank you for your attention!!