Relating Label Semantics and Prototype Theory

Jonathan Lawry and Yongchuan Tang

University of Bristol and Zhejiang University

Outline

- Prototype theory of concepts.
- Prototype interpretation of fuzziness.
- Random set interpretation of fuzziness.
- Background to label semantics.
- A prototype based interpretation of appropriateness measures.
- Conclusions.

Concept Representation

- In logic (e.g. Kripke Semantics) concepts are represented as mappings from a set of possible worlds into sets of element, each corresponding to a possible extension of the concept.
- This fails to take account of the role similarity plays in establishing the meaning of concept labels.
- In Prototype theory (Rosch 1975) concepts are defined in terms of similarity to prototypical points (in a conceptual space).

Prototype Interpretation of Fuzzy Sets

- Prototype similarity has been proposed as an interpretation of fuzzy membership functions.
- A similarity measure S is defined on the underlying domain, taking values in $[0,1]$.
- Ruspini (1991): Membership of x in L_i corresponds to $\sup\{S(x,y) : y \in P_i\}$, where P_i is the set of prototypes of L_i.
- Resulting calculus is not truth-functional.

Random Set Interpretation of Fuzzy Sets

- Seminal work of Nguyen and Goodman proposed a random set interpretation of fuzzy sets.
- The extensions of vague concepts are represented by random sets of the underlying domain characterising uncertain boundaries.
- Membership functions then correspond to single point coverage functions of random sets; i.e. the probability that an element is a member of the random set.
- Resulting calculus is not truth-functional.
- Labels semantics is also a random set model but where the random sets are on labels rather than values.
- This new work links random set and prototype theory.

The Label Semantics Framework

- Label Semantics provides an alternative model of the uncertainty associated with the use of vague linguistic labels to describe objects/values.
- The basis of the model is in judging the appropriateness or assertibility of vague expressions to describe an object or an instance.
- Appropriateness in this context is governed by the conventions of language use emergent from a population of communicating agents.
- An agent’s knowledge of appropriateness is based on partial evidence concerning the previous use of labels and is consequently uncertain.

The Epistemic Stance

- Label semantics (and other random set models) assumes that communicating agents adopt an epistemic stance regarding labels.
- *Within a population of communicating agents, individuals assume the existence of a set of labelling conventions for the population governing what linguistic labels and expression can be appropriately used to describe particular instances.*
- Making an assertion to describe an object or instance x involves making a decision as to what labels can be appropriately used to describe x.

Measures of Appropriateness

- Assume that there is a finite set of labels $LA = \{L_1, \ldots, L_n\}$ for describing elements of the universe Ω.
- LE is the set of expressions generated from LA through recursive application of the connectives $\wedge, \vee$ and $\neg$.
- For $\theta \in LE$, $x \in \Omega$, $\mu_\theta(x) =$ the subjective probability that θ is appropriate to describe x.

Mass Functions

- $\mathcal{D}_x$ is the complete set of labels appropriate to describe x.
- $\mathcal{D}_x$ is a random set into 2^{LA}.
- $m_x : 2^{LA} \rightarrow [0, 1]$ is a probability mass function on subsets of labels.
- For $S \subseteq LA$ $m_x(S)$ is the subjective probability that $\mathcal{D}_x = S$.
- The mass function m_x and the appropriateness measure μ are strongly related...
- $\mu_\theta(x)$ is the sum of m_x over those values for $\mathcal{D}_x$ consistent with θ

General Relationships

'x is θ' requires that $\mathcal{D}_x \in \lambda(\theta)$ where $\forall \theta, \varphi \in LE$

- $\forall L \in LA\ \lambda(L) = \{S \subseteq LA : L \in S\}\ (= \{S \in \mathcal{F} : L \in S\})$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$, $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$

This results in the following equation relating m_x as μ:

- $\mu_\theta(x) = P(\mathcal{D}_x \in \lambda(\theta)) = \sum_{S \in \lambda(\theta)} m_x(S)$

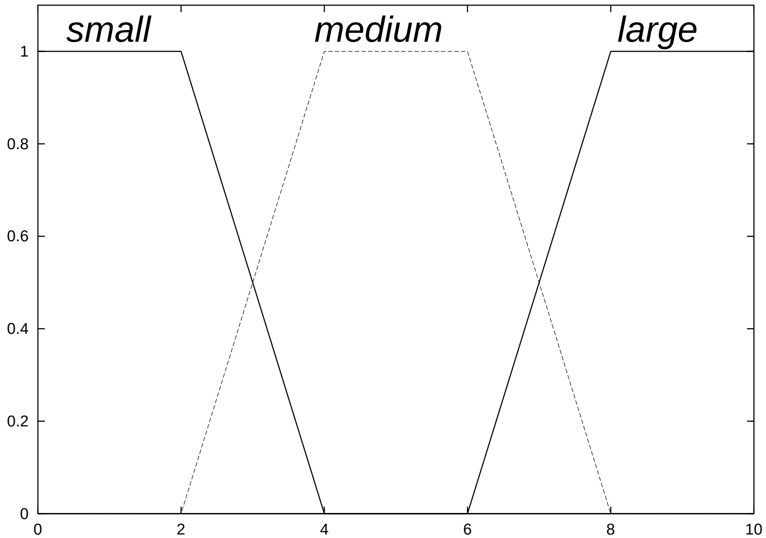
Functionality:

- By functionality we mean that for any θ there is a function $f_\theta : [0,1]^n \rightarrow [0,1]$ such that $\forall x$ $\mu_\theta(x) = f_\theta(\mu_{L_1}(x), \ldots \mu_{L_n}(x))$.
- Appropriateness measures are not in general functional except under certain conditions ...

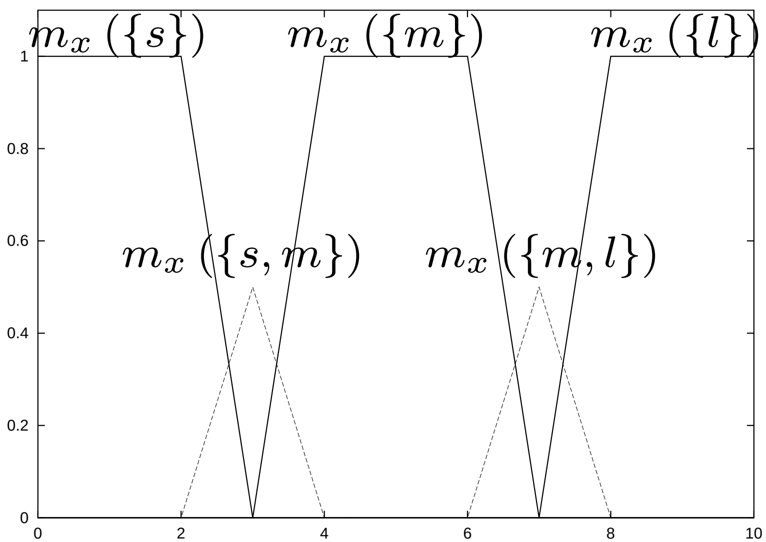
Ordering on Labels

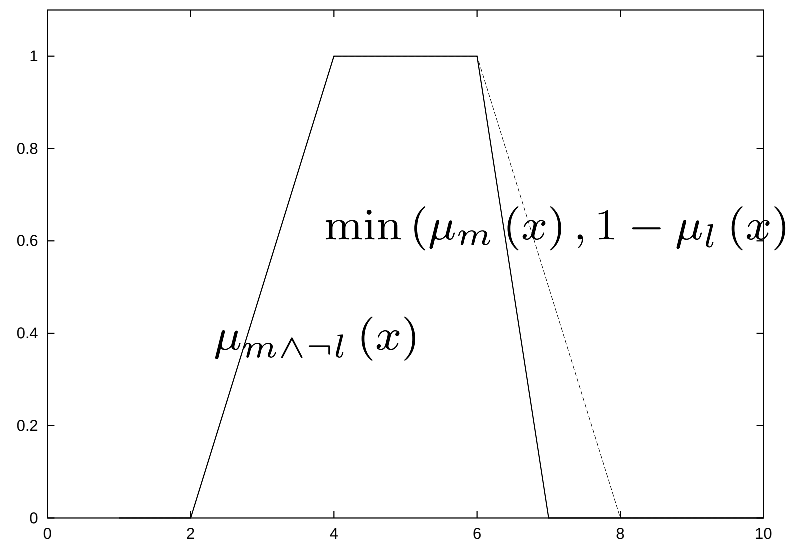
- Agents rank the labels LA in terms of their appropriateness according to the ordering $\preceq_x$, so that $L \preceq_x L'$ means that L' is judged at least as appropriate as L for describing x.
- If $L \preceq_x L'$ then $L \in \mathcal{D}_x$ implies that $L' \in \mathcal{D}_x$.
- If $\preceq_x$ is a total ordering then possible values of $\mathcal{D}_x$ form a nested hierarchy i.e. $\mathcal{D}_x$ is a *consonant* random set.
- If $\preceq_x$ is a partial ordering then...

The Consonance Assumption



$\mu_{L_1}(x), \ldots, \mu_{L_n}(x)$ ordered so that
$\mu_{L_i}(x) \geq \mu_{L_{i+1}}(x)$ for $i = 1, \ldots, n-1$
$m_x(\{L_1, \ldots, L_n\}) = \mu_{L_n}(x)$
$m_x(\{L_1, \ldots, L_i\}) = \mu_{L_i}(x) - \mu_{L_{i+1}}(x)$
and $m_x(\emptyset) = 1 - \mu_{L_1}(x)$





Properties under Consonance

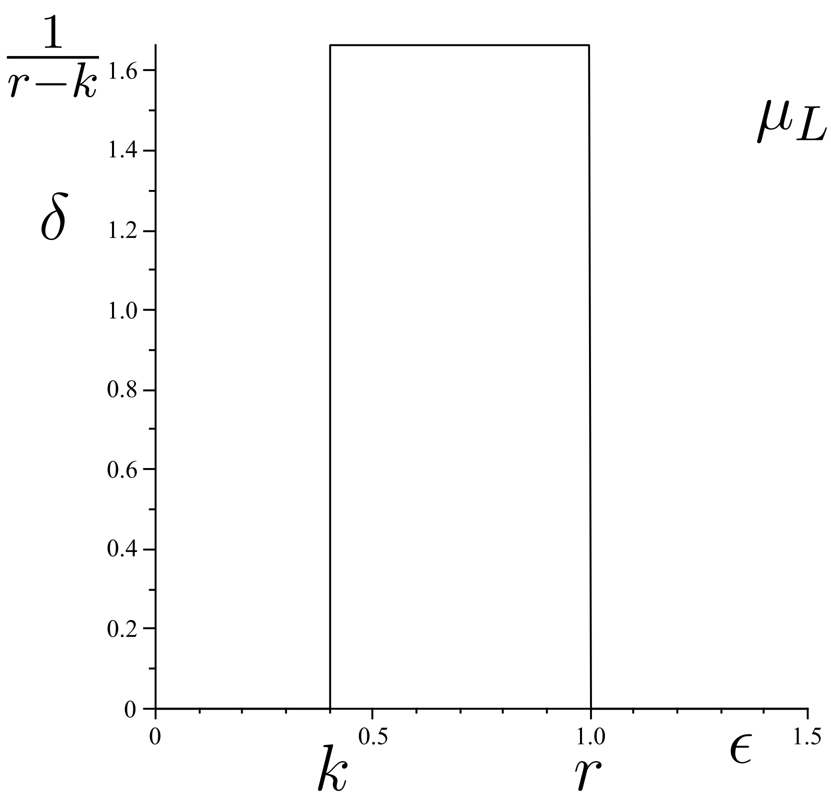
- If $\models \theta$ then $\forall x,\ \mu_\theta(x) = 1$
- If $\theta \equiv \varphi$ then $\forall x,\ \mu_\theta(x) = \mu_\varphi(x)$
- $\forall \theta, \forall x,\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$
- $\forall \theta, \varphi \in LE^{\wedge,\vee}$, $\forall x \in \Omega$ it holds that $\mu_{\theta\wedge\varphi}(x) = \min(\mu_\theta(x), \mu_\varphi(x))$ and $\mu_{\theta\vee\varphi}(x) = \max(\mu_\theta(x), \mu_\varphi(x))$
- Law of Excluded Middle: $\lambda(\theta) \cup \lambda(\theta)^c = 2^{2^{LA}}$ therefore $\forall x\ \mu_{\theta\vee\neg\theta}(x) = 1$
- Law of Non-contradiction: $\lambda(\theta) \cup \lambda(\theta)^c = \emptyset$ therefore $\forall x\ \mu_{\theta\wedge\neg\theta}(x) = 0$

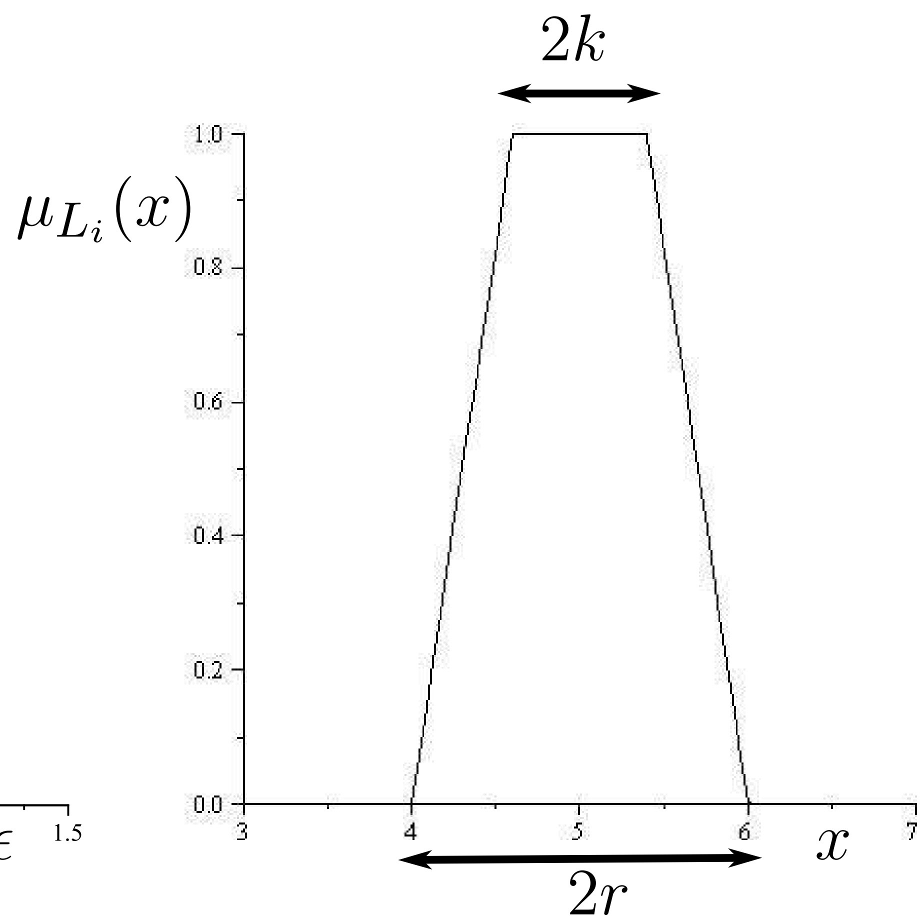
A Prototype Interpretation

- Let $d : \Omega^2 \rightarrow [0, \infty)$ be a distance metric satisfying $d(x, x) = 0$ and $d(x, y) = d(y, x)$.
- For $L_i \in \Omega$ let $P_i \subseteq$ be a set of prototypical elements.
- For $x \in \Omega$ let $d(x, P_i) = \sup\{d(x, y) : y \in P_i\}$
- Let ϵ be a random variable into $[0, \infty)$ with density function δ.
- L_i is appropriate to describe x iff $d(x, P_i) \leq \epsilon$.
- $\mathcal{D}_x^\epsilon = \{L_i : d(x, P_i) \leq \epsilon\}$.
- $\forall F \subseteq \Omega \; m_x(F) = \delta(\{\epsilon : \mathcal{D}_x^\epsilon = F\})$
- $L_i \preceq_x L_j$ iff $d(x, P_i) \geq d(x, P_j)$ (a total ordering)

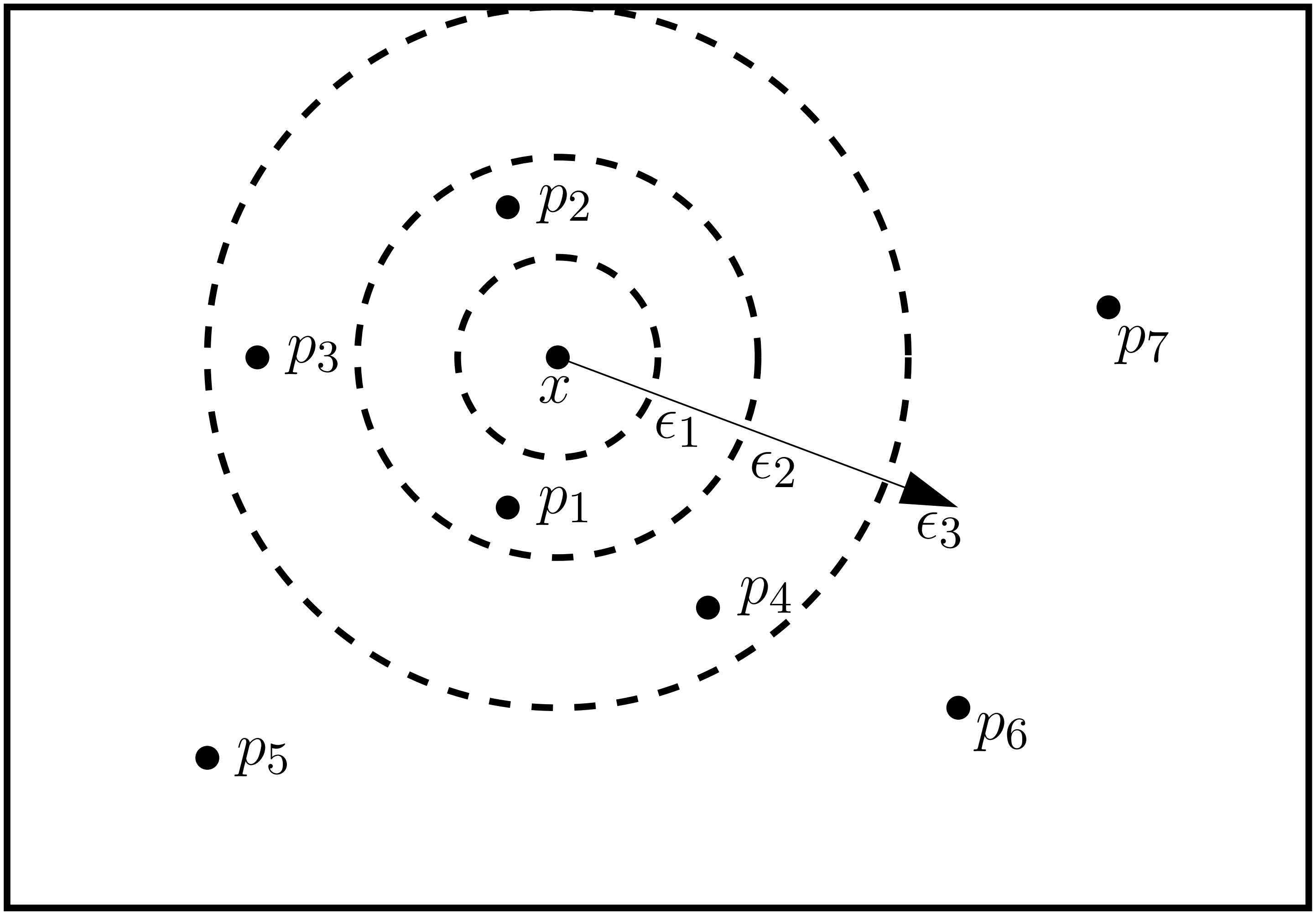
Appropriateness of Basic Labels

- $\mu_{L_i}(x) = \delta(\{\epsilon : d(x, P_i) \leq \epsilon\}) = \delta([d(x, P_i), \infty))$.
- Example $L_i = about\ 5$ and $P_i = \{5\}$





Generating $\mathcal{D}_x^\epsilon$



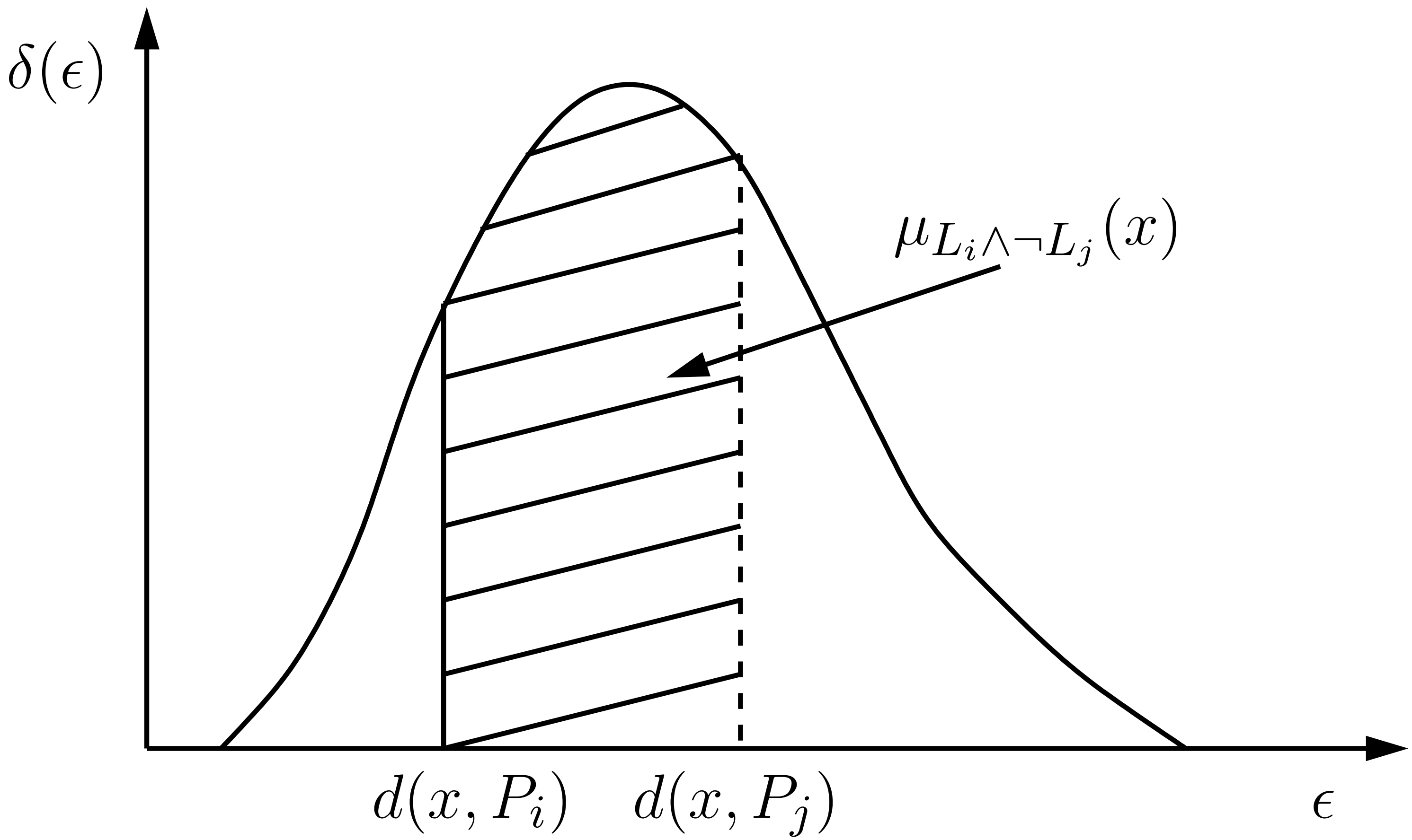
$P_i = \{p_i\} : i = 1, \ldots, n$: Identifying $\mathcal{D}_x^\epsilon$; $\mathcal{D}_x^{\epsilon 1} = \emptyset$, $\mathcal{D}_x^{\epsilon 2} = \{L_1, L_2\}$, $\mathcal{D}_x^{\epsilon 3} = \{L_1, L_2, L_3, L_4\}$

Appropriateness as Intervals of ϵ

- $\forall x \in \Omega$ and $\forall \theta \in LE$, $I(\theta, x) \subseteq [0, \infty)$ is defined recursively by:
- $\forall L_i \in LA\ I(L_i, x) = [d(x, P_i), \infty)$
- $\forall \theta \in LE\ I(\neg\theta, x) = I(\theta, x)^c$
- $\forall \theta, \varphi \in LE\ I(\theta \vee \varphi, x) = I(\theta, x) \cup I(\varphi, x)$
- $\forall \theta, \varphi \in LE\ I(\theta \wedge \varphi, x) = I(\theta, x) \cap I(\varphi, x)$
- Theorem: $\forall \theta \in LE, \forall x \in \Omega\ I(\theta, x) = \{\epsilon : \mathcal{D}_x^\epsilon \in \lambda(\theta)\}$
- Corollary: $\forall \theta \in LE, \forall x \in \Omega\ \mu_\theta(x) = \delta(I(\theta, x))$

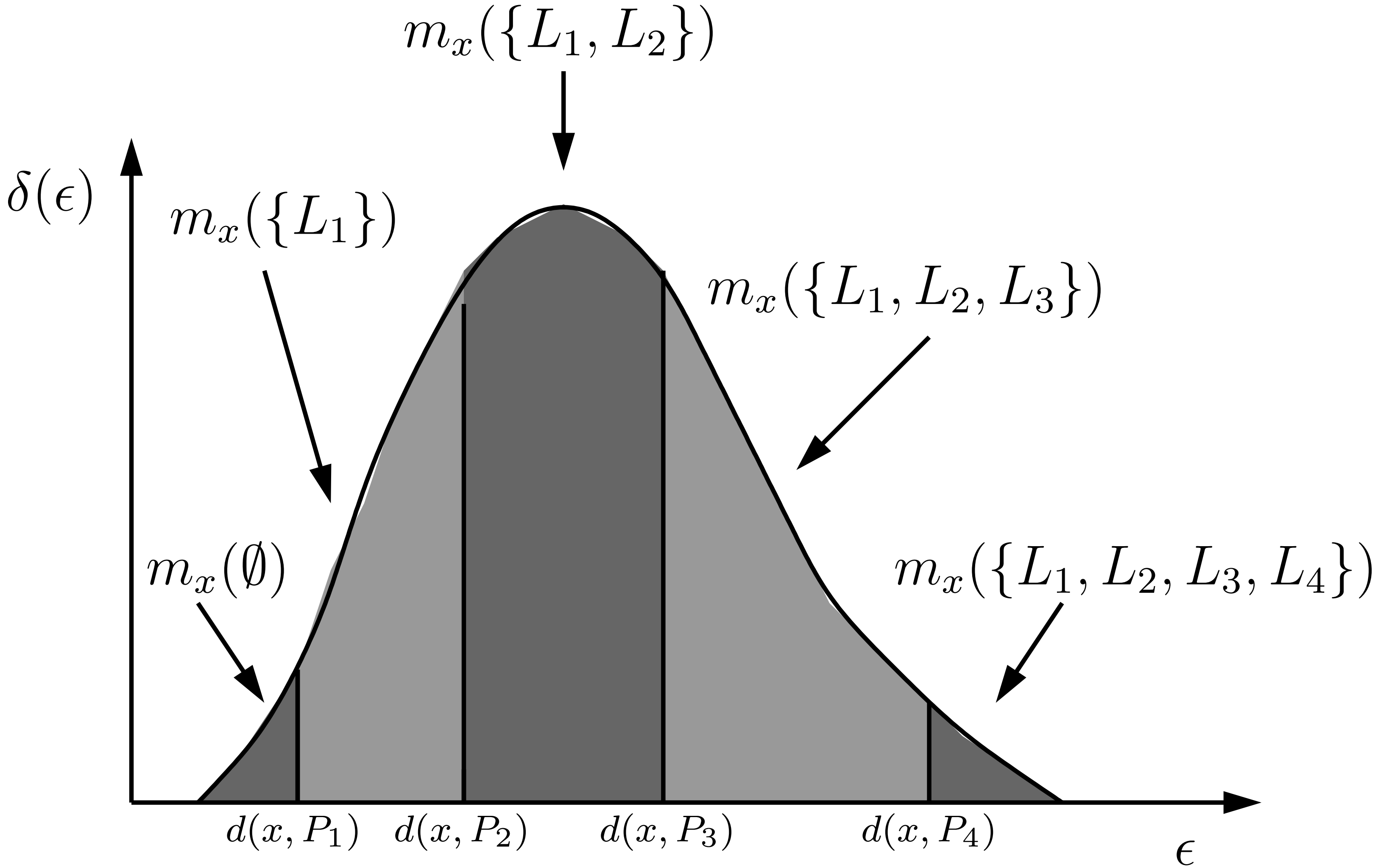
Area under δ: Appropriateness

$$I(L_i \wedge \neg L_j, x) = [d(x, P_i), d(x, P_j))$$



Area under δ: Mass

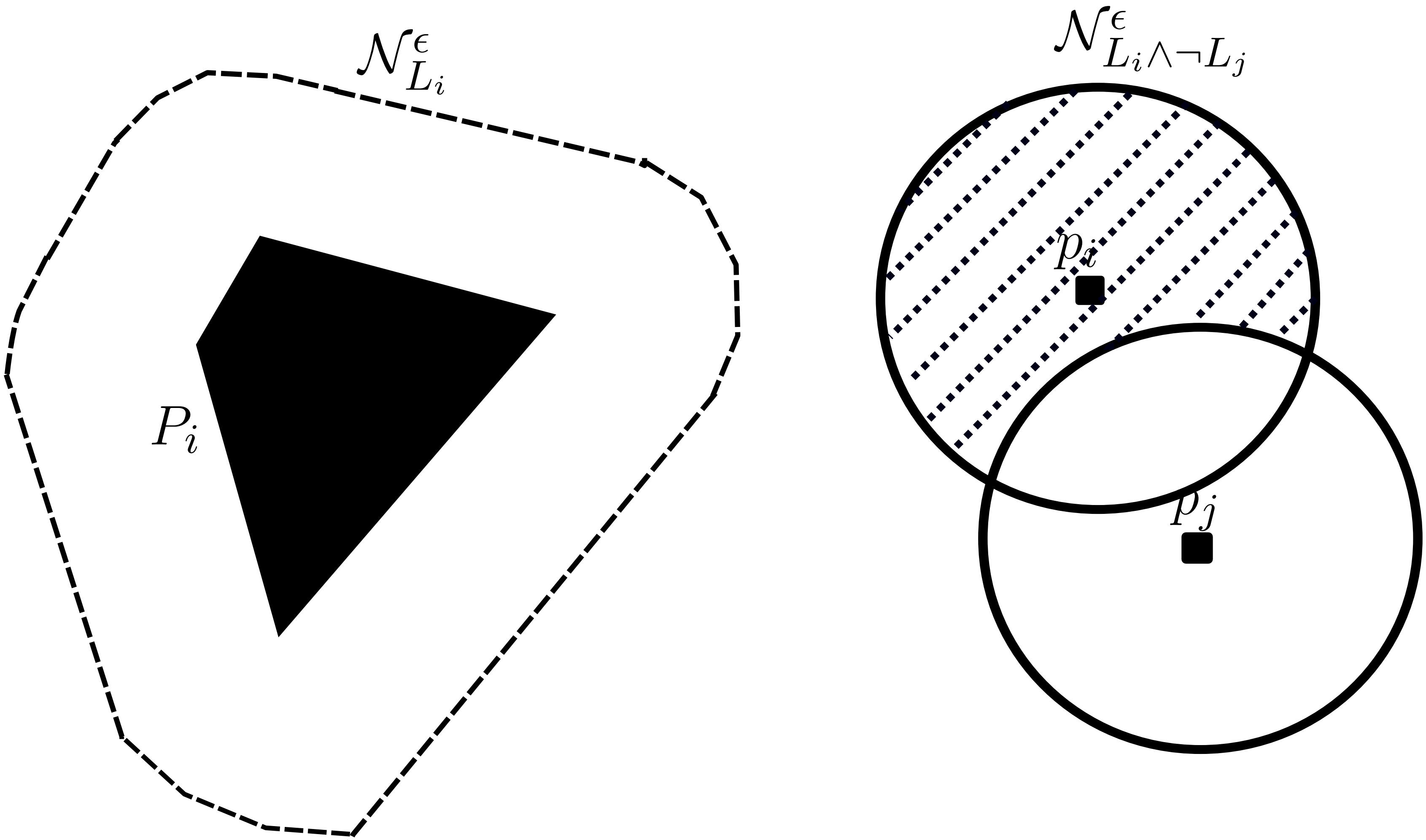
$$m_x(F) = \delta([\max\{d(x,P_i) : L_i \in F\}, \min\{d(x,P_i) : L_i \notin F\}))$$



Another Perspective

- Consider a neighbourhood of label prototypes P_i: $\mathcal{N}^{\epsilon}_{L_i} = \{x : d(x, P_i) \leq \epsilon\} = \{x : L_i \in \mathcal{D}^{\epsilon}_x\}$ $= \{x : L_i$ is appropriate to describe $x\}$.
- The ϵ-extension of the concept L_i is the set of points ϵ-similar to the prototype(s) of L_i.
- Extended to any expression θ: $\mathcal{N}^{\epsilon}_{\theta} = \{x : \mathcal{D}^{\epsilon}_x \in \lambda(\theta)\} = \{x : \epsilon \in I(\theta, x)\}$
- Appropriateness measures: (Theorem) $\mu_\theta(x) = P(x \in \mathcal{N}^{\epsilon}_{\theta}) = \delta(\{\epsilon : x \in \mathcal{N}^{\epsilon}_{\theta}\})$
- Viewing $\mathcal{N}^{\epsilon}_{\theta}$ as a random set then $\mu_\theta(x)$ is the single point coverage function for $\mathcal{N}^{\epsilon}_{\theta}$.
- If $\theta \in LE^{\wedge,\vee}$ then $\forall \epsilon \leq \epsilon' \; \mathcal{N}^{\epsilon}_{\theta} \subseteq \mathcal{N}^{\epsilon'}_{\theta}$.

Prototype Neighbourhoods



Recursive Definition of $\mathcal{N}_\theta^\epsilon$

- The neighbourhoods $\mathcal{N}_\theta^\epsilon$ can also be defined directly in a recursive manner:
- For $L_i \in LA$ $\mathcal{N}_{L_i}^\epsilon = \{x \in \Omega : d(x, P_i) \leq \epsilon\}$
- For $\theta, \varphi \in LE$ $\mathcal{N}_{\theta \wedge \varphi}^\epsilon = \mathcal{N}_\theta^\epsilon \cap \mathcal{N}_\varphi^\epsilon$
- For $\theta, \varphi \in LE$ $\mathcal{N}_{\theta \vee \varphi}^\epsilon = \mathcal{N}_\theta^\epsilon \cup \mathcal{N}_\varphi^\epsilon$
- For $\theta \in LE$ $\mathcal{N}_{\neg\theta}^\epsilon = (\mathcal{N}_\theta^\epsilon)^c$
- Hence, appropriateness measures can be viewed as single point coverage functions of random sets into 2^Ω, generated recursively from neighbourhoods of prototypes according to standard boolean combination rules.

A General Model

- For each label L_i we define a distinct metric $d_i : \Omega \rightarrow [0, \infty)$ and threshold ϵ_i.
- $\delta(\epsilon_1, \ldots, \epsilon_n)$ is then the joint density on all threshold variables.
- $\mathcal{D}_x^{\vec{\epsilon}} = \{L_i : d_i(x, P_i) \leq \epsilon_i\}$ and $\mu_\theta(x) = \delta(\{\vec{\epsilon} : \mathcal{D}_x^{\vec{\epsilon}} \in \lambda(\theta)\})$
- Independence: $\delta(\epsilon_1, \ldots, \epsilon_n) = \prod_{i=1}^n \delta_i(\epsilon_i)$ and $m_x(F) = \prod_{L_i \in F} \mu_{L_i}(x) \times \prod_{L_i \notin F}(1 - \mu_{L_i}(x))$
- Consonance: $\epsilon_i = f_i(\epsilon)$ where $f_i : [0, \infty) \rightarrow [0, \infty)$ is an increasing function.

Conclusion

- A prototype interpretation of label semantics has been introduced.
- Appropriateness measures are interpreted as the probability that a distance threshold ϵ lies with a particular interval of $[0, \infty)$ as determined by the relevant expression.
- ϵ is a random variable representing the upper bound on $d(x, P_i)$ at which L_i can still be deemed appropriate to describe x.
- Appropriateness measures can also be defined in terms of random set neighbourhoods of prototypes. i.e. $L_i =$ approximately P_i