Linguistic Approximation: Outline

- Definition and Problem Formulation
- Motivation
  - – What's wrong with distances in the membership space?
- The Engineering Solution: the PR Approach
  - – Feature Extraction
    - Feature Selection
  - – Classification (Distances and Spaces)
    - Weights Determination (Saaty)
    - Refinements
- Reformulation for parametrized membership funct.
- Example of Applications

Linguistic Approximation: Definition

- What is it?

  - The process of *labeling* a membership distribution with a symbol whose meaning is *the closest* (according to some distance) to the meaning of the unlabelled distribution

  - In other words, we associate a label with a distribution on the basis of semantic similarity

  - Formally it is a mapping from the crisp set of all fuzzy subsets of the universe of discourse U into the language ℓ=L(G_1): $LA : U \rightarrow L(G_1)$

Linguistic Approximation: Purpose

- Why do we need it?
  - To force the closure within a language of results of operations that do not preserve such closure
  - To simplify existing labels
  - To provide a linguistic summary of the result of an external process (approximate reasoning, extension principle, etc.)

Problem Formulation

- The Linguistic Approximation of an unlabelled fuzzy set A'(u) is:

$$LA[A'(u)] = S_A$$

such that

$$d\Big(A'(u), A(u)\Big) = \min_{\forall A_i \in L(G_1)} d\Big(A'(u), A_i(u)\Big)$$

Problem Formulation

where:

- A and A' are subset of the same universe U
- A' is the unlabelled fuzzy set
- S_A is the label (sentence) corresponding to A(u)
- $L(G_1)$ is the finite termset (target language)
- d(.,.) is a distance satisfying the axioms of a metric, i.e.:

$$d(A,B) > 0 \leftrightarrow (A \neq B)$$

$$d(A,B) = 0 \leftrightarrow (A = B)$$

$$d(A,B) = d(B,A)$$

$$d(A,C) \leq d(A,B) + d(B,C)$$

Problem with This Formulation

- Dependence on Distance Selection
  - Hamming Distance (First Norm) $\| \, . \, \|_1$

$$d_1\big(A(u_k), B(u_k)\big) = \frac{1}{D}\sum_{k=1}^{D} |A(u_k) - B(u_k)|$$

  - Euclidean Distance (Second Norm) $\| \, . \, \|_2$

$$d_2\big(A(u_k), B(u_k)\big) = \left[\frac{1}{D}\sum_{k=1}^{D}\big(A(u_k) - B(u_k)\big)^2\right]^{0.5}$$

  - Max-based Distance (Infinite Norm) $\| \, . \, \|_\infty$

$$d_\infty\big(A(u_k), B(u_k)\big) = \max_{k=1}^{D}\big(A(u_k) - B(u_k)\big)$$

Problem with This Formulation

- Max-based distance
  - Is too sensitive to a mismatch in one point
  - Saturates (d=1) very quickly - see Excel File

$$d_{inf}\ (Low, More\_or\_less\ Low) = 0.246$$

$$d_{inf}\ (Low, Sort\_of\ Low) = d_{inf}\ (Low, Middle) = d_{inf}\ (Low, High) = 1$$

- Hamming and Euclidean Distances
  - When the distributions are disjoints, the distances reflect the sum the areas under the two distributions
  - As a result we may have inconsistent distances:

$$d_1\ (Low, High) > d_1 (Very\_Low,\ Very\_High)$$

  - It does not take uniform values within [0,1] - only on a limited range of values

Example

Universe	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Low	1	0.98	0.92	0.82	0.68	0.5	0.32	0.18	0.08	0.02	0	0	0	0	0	0	0	0	0	0	0
Middle	0	0	0	0	0	0.06	0.22	0.5	0.78	0.94	1	0.94	0.78	0.5	0.22	0.06	0	0	0	0	0
High	0	0	0	0	0	0	0	0	0	0	0	0.02	0.18	0.32	0.5	0.68	0.68	0.82	0.92	0.98	1
All	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
None	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Connectors and Modifiers

Very (X) = X^2

More_or_less (X) = $X^{0.5}$

Sort_of (X) = Norm[Not Very Very (X) And More_or_Less (X)]

X And Y = Min (X,Y)

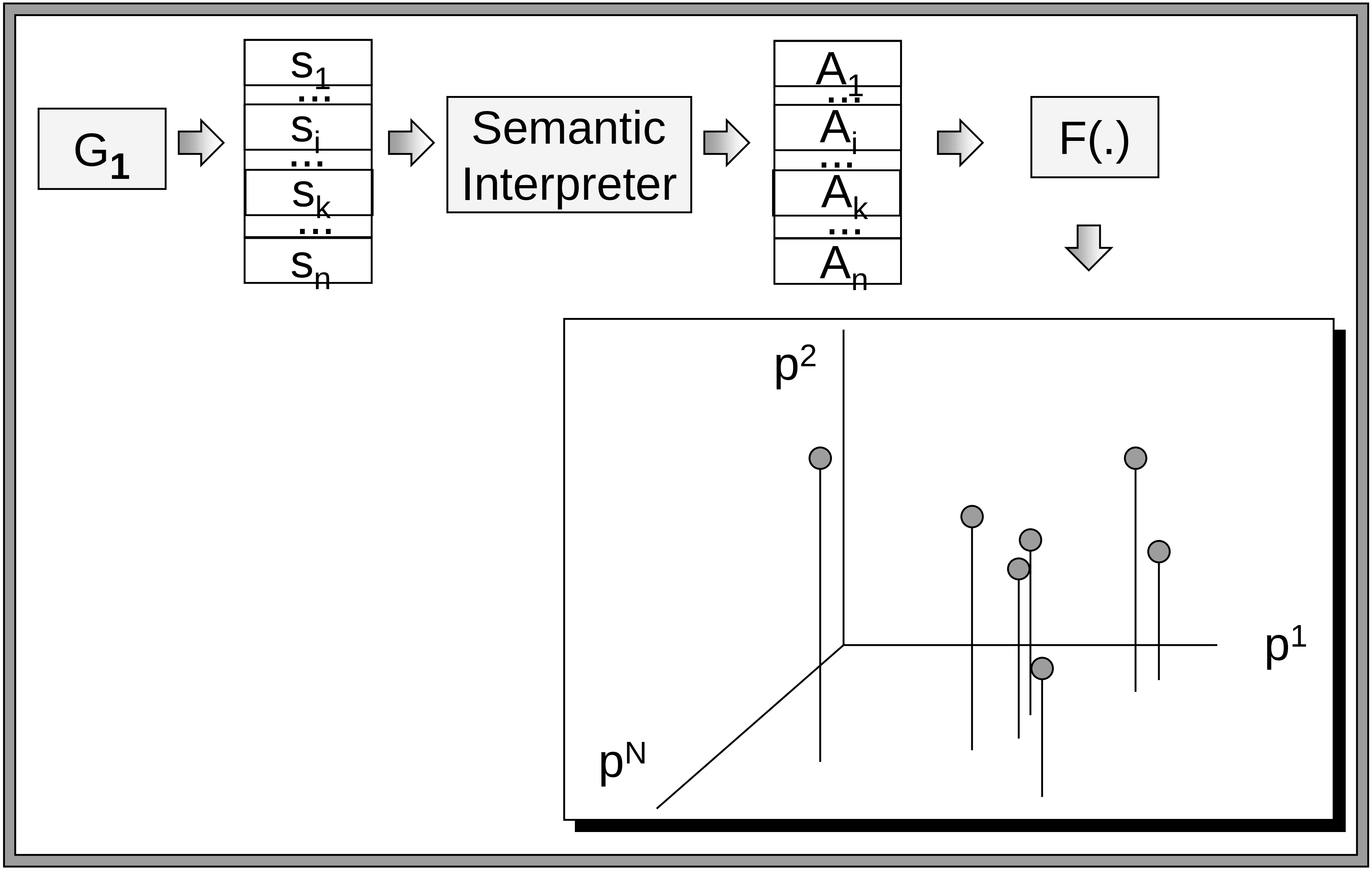
X Or Y = Max (X,Y)

Not (X) = 1-X

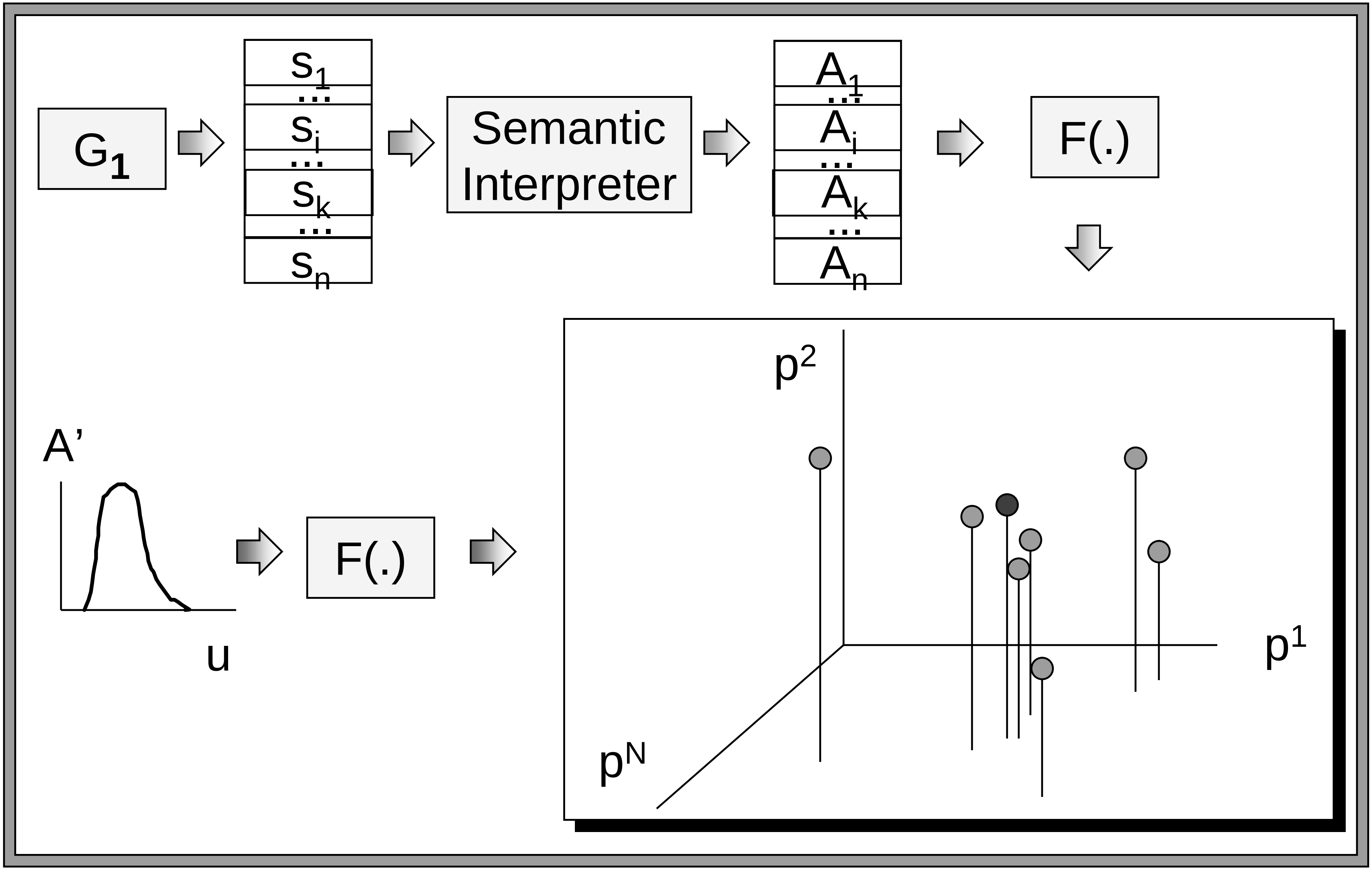
Example (cont)

	d1	d2	dinf
| Low - Very Low |	0.06	0.10	0.25
| Low - More_or_less Low |	0.06	0.11	0.25
| Low - Sort_of Low |	0.21	0.36	1
| Low - More_or_less Middle |	0.54	0.67	1
| Low - Middle |	0.49	0.63	1
| Low - High |	0.55	0.66	1
| Low - Very High |	0.49	0.61	1
| Very Low - Very High |	0.43	0.58	1

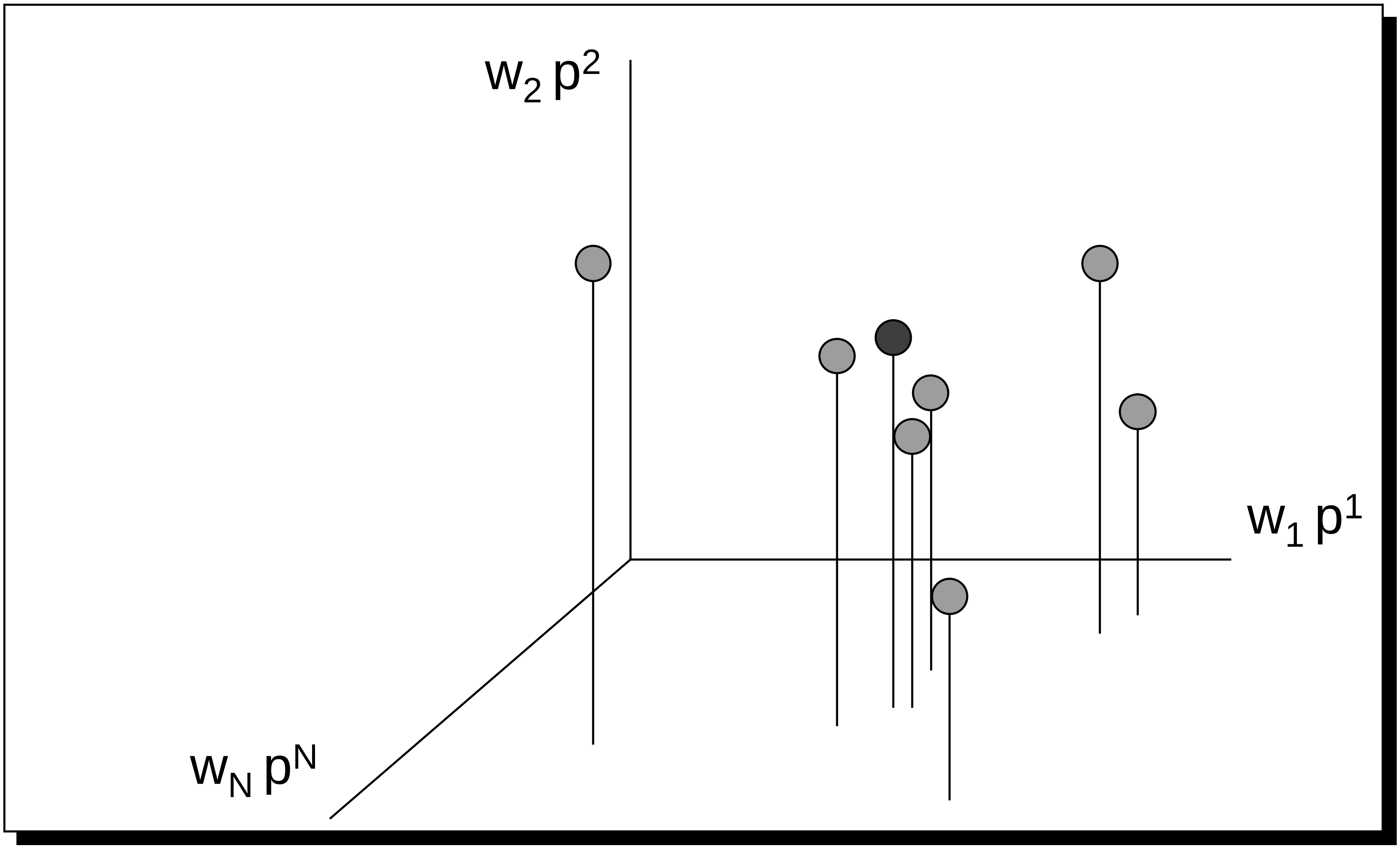
The Engineering Solution: PR Approach (1)



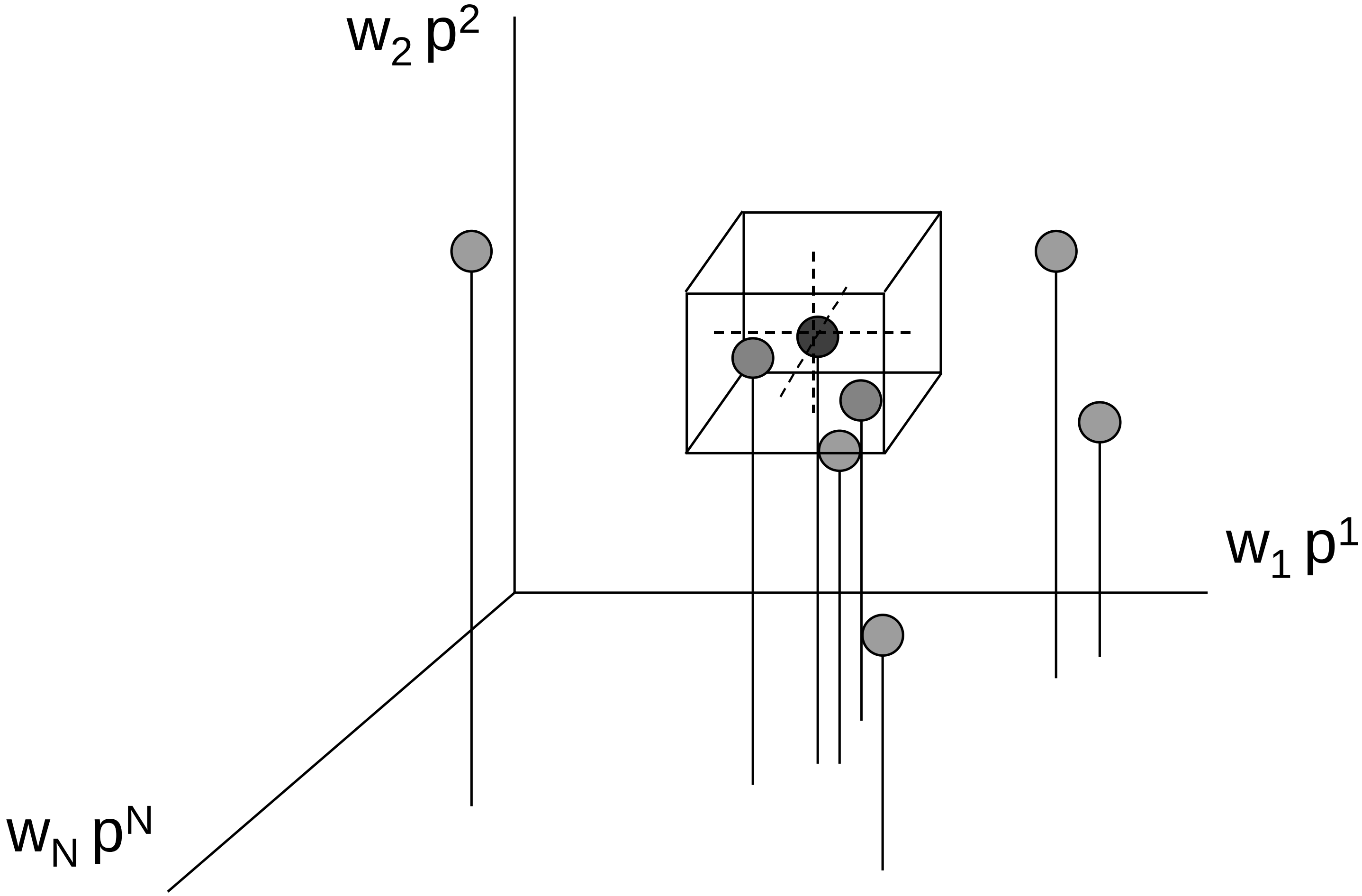
The Engineering Solution: PR Approach (2)



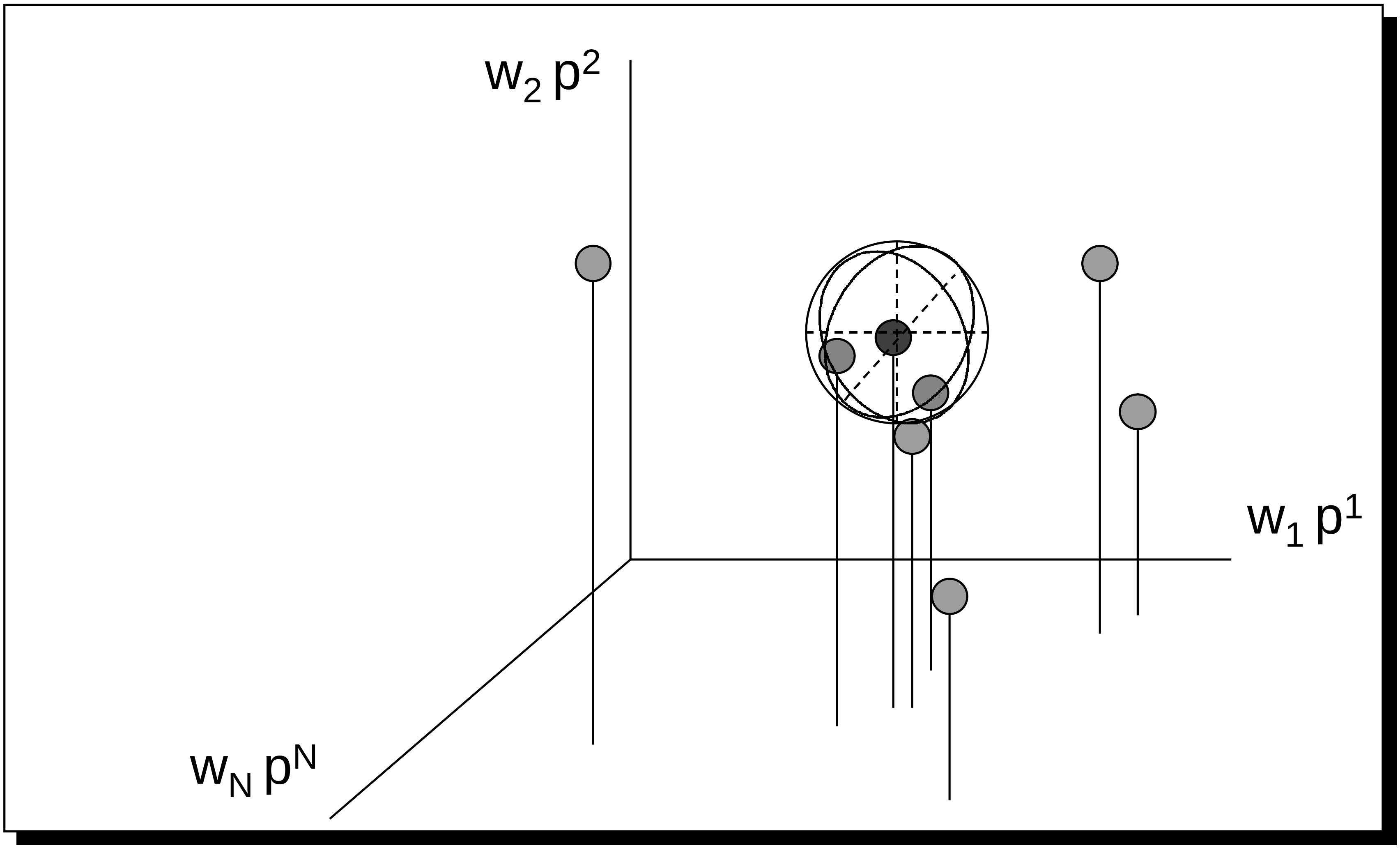
The Engineering Solution: PR Approach (3)



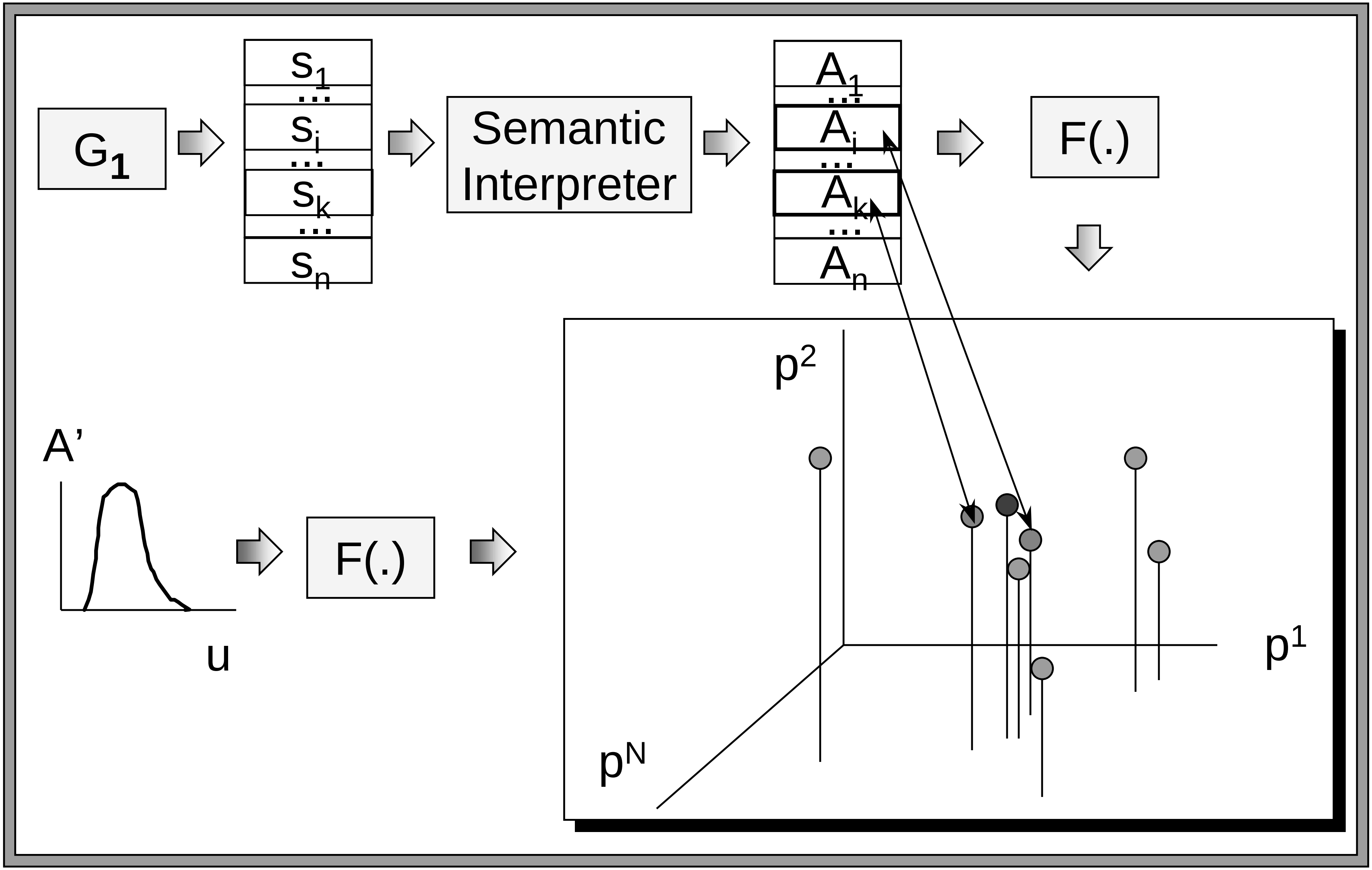
The Engineering Solution: PR Approach (4)



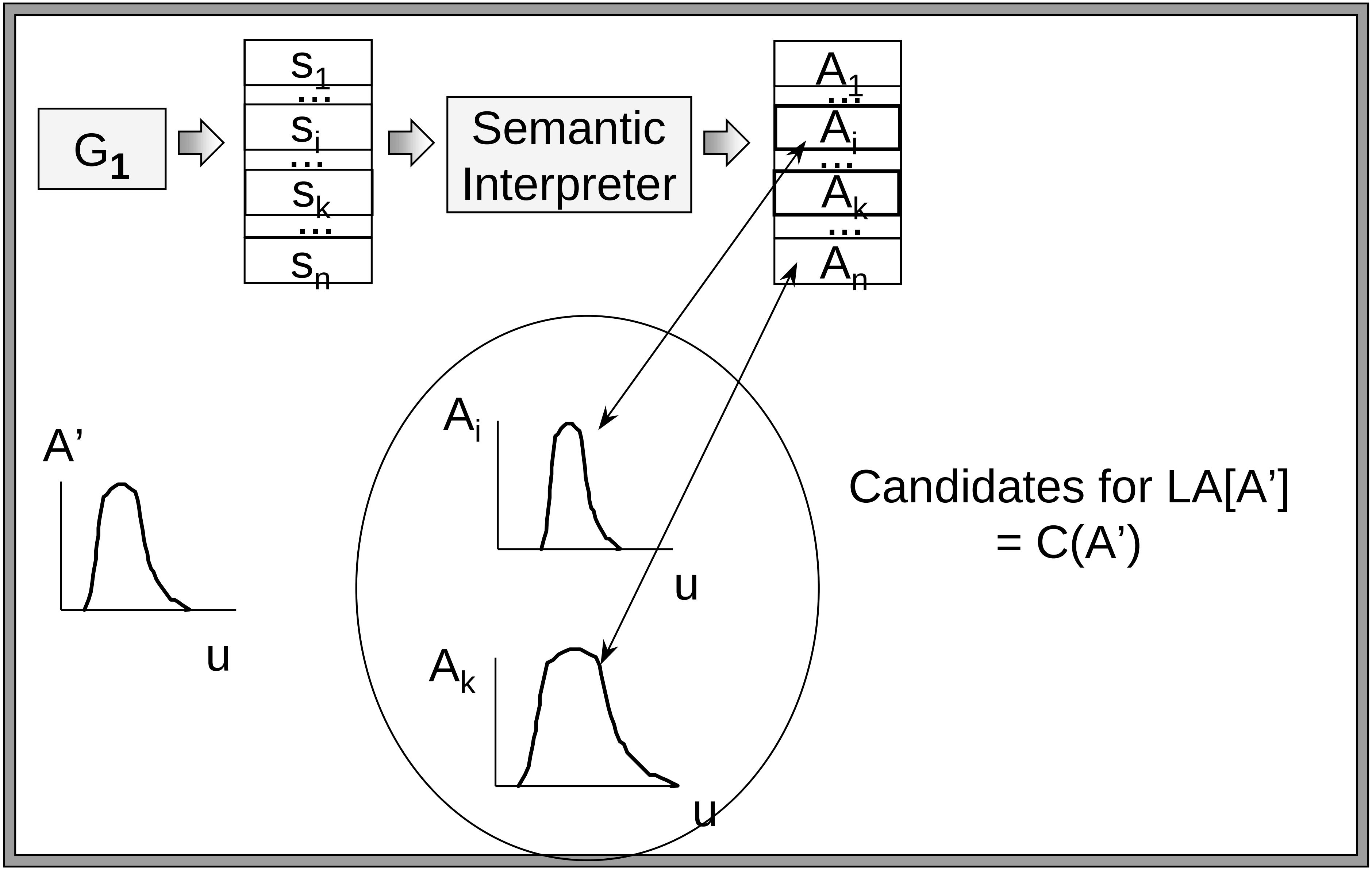
The Engineering Solution: PR Approach (4')



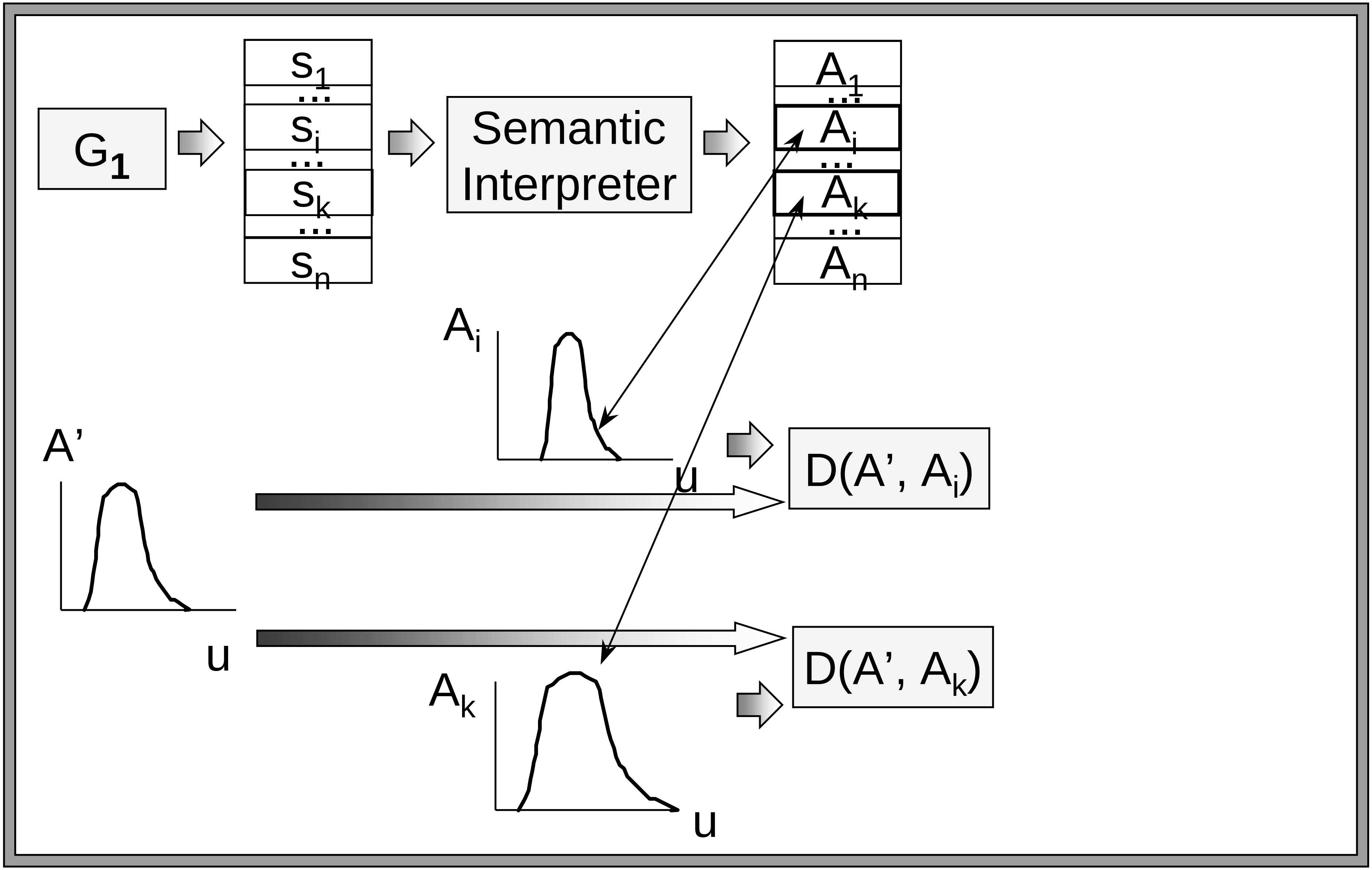
The Engineering Solution: PR Approach (5)



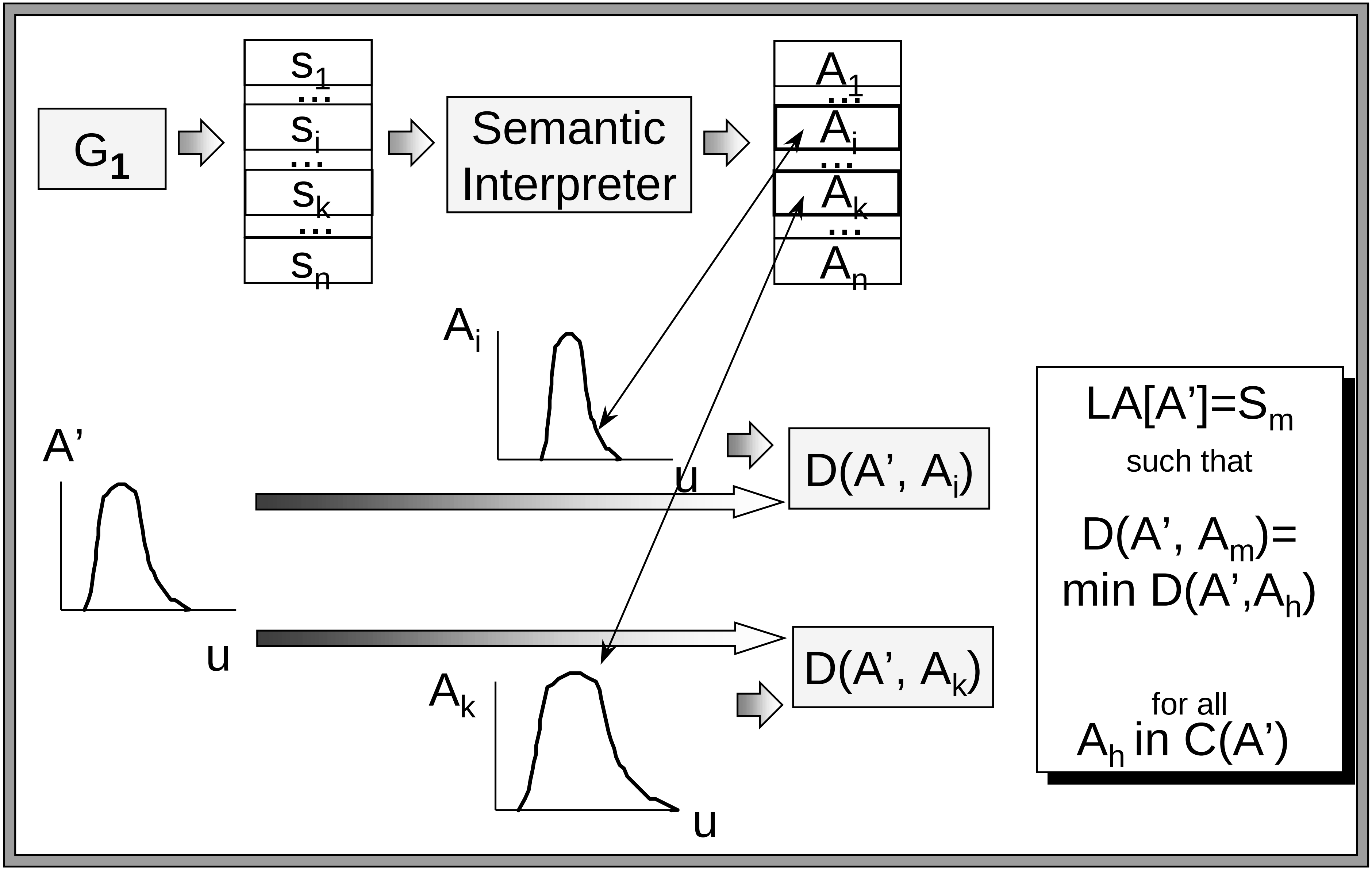
The Engineering Solution: PR Approach (6)



The Engineering Solution: PR Approach (7)



The Engineering Solution: PR Approach (8)



Feature Extraction and Classification

- Starting from the assumption that the target language is finite:

$$|L(G_1)| = n < \infty$$

- and that the universe of discourse U is sampled into D points, i.e. |U| = D

- we define a mapping F:

$$F : [0,1]^D \to R^N, \text{ where } N << D$$

Feature Extraction (cont.)

- F maps any fuzzy set $A_i(u)$ into the N dimensional feature space P
- Each element in P is a point (denoted by the vector $\vec{P}_i$ corresponding to the values of the features of the membership function $A_i(u)$

$$F\left(A_i(u)\right) = \vec{P}_i = (p_i^1, p_i^2, \ldots, p_i^N)$$

- After several experiments we decided to use the following four features (i.e., N = 4):

Features Selection

- Sigma Cardinality (Power)

$$p_i^1(A_i) = \sum_{k=1}^{D} A_i(u_k)$$

- Measure of Fuzziness

$$p_i^2(A_i) = \left\| H\big(A_i(u_k)\big) \right\|_2 = \sum_{k=1}^{D} \left[H^2\big(A_i(u_k)\big) \right]^{0.5}$$

where:

$$H(x) = \begin{cases} x & \text{if } x < 0.5 \\ (1-x) & \text{if } x \geq 0.5 \end{cases}$$

Features Selection

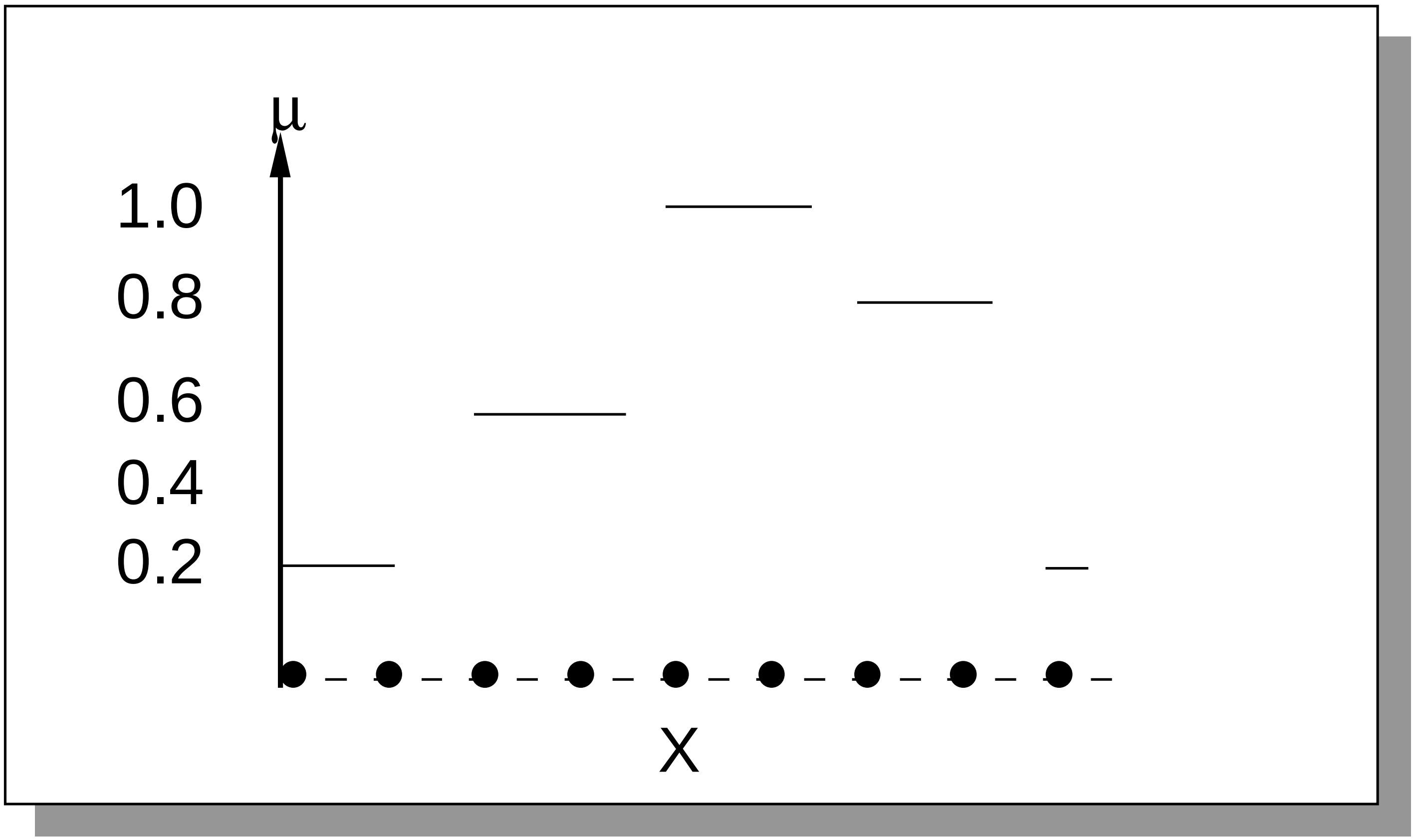
- First Moment (Center of Gravity)

$$p_i^3(A_i) = \hat{u} = \frac{\sum_{k=1}^{D} u_k \times A_i(u_k)}{\sum_{k=1}^{D} A_i(u_k)} = \sum_{k=1}^{D} u_k \times \frac{A_i(u_k)}{p_i^1(A_i)}$$

- Third Moment (Skewness)

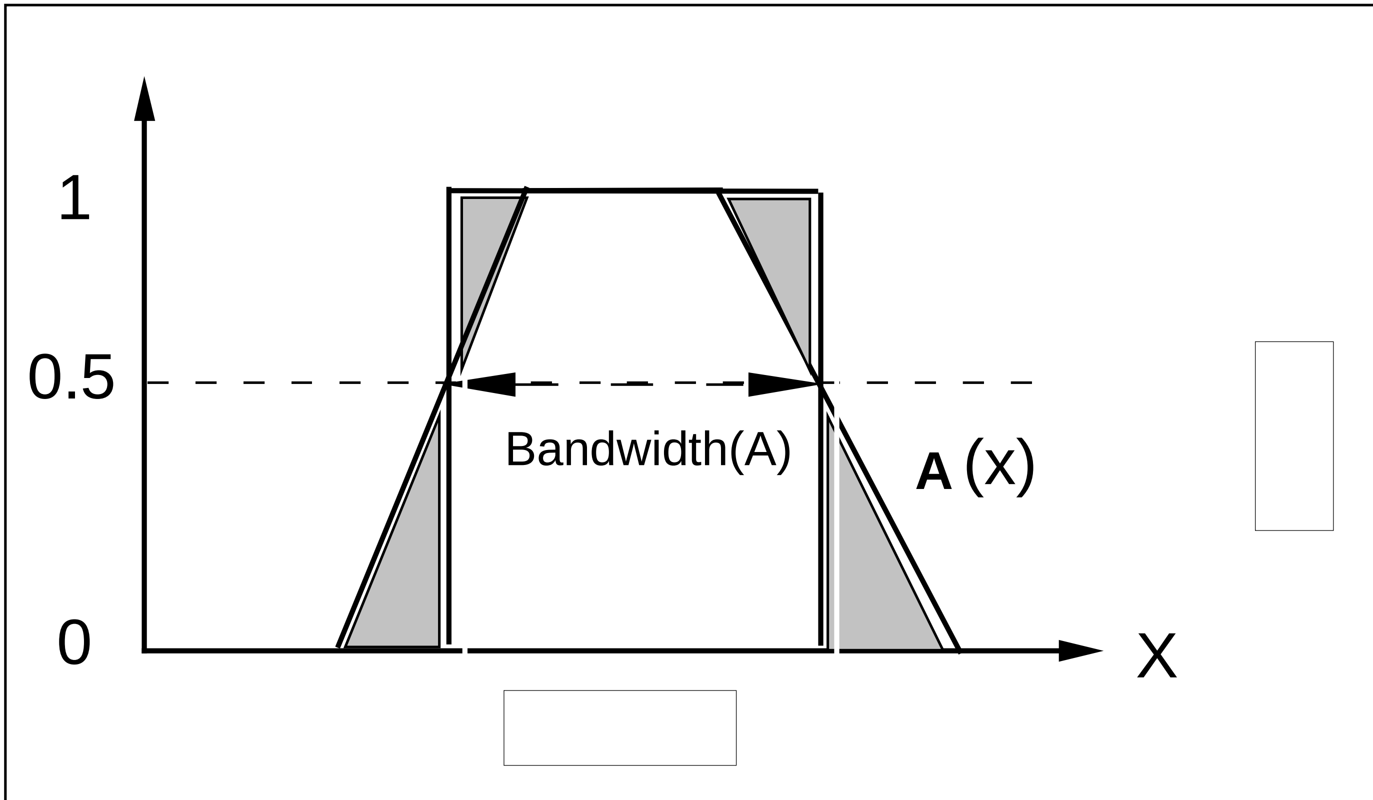
$$p_i^4(A_i) = \frac{\sum_{k=1}^{D} (u_k - \hat{u})^3 \times A_i(u_k)}{\sum_{k=1}^{D} A_i(u_k)} = \sum_{k=1}^{D} (u_k - \hat{u})^3 \times \frac{A_i(u_k)}{p_i^1(A_i)}$$

Sigma Cardinality



Σ **Count (A) = Card(A) = 5.4**

Measure of Fuzziness



Measure of Fuzziness = Cardinality {|Bandwidth(A)- A(x)|}
= Cardinality { ▭ } = || ▭ ||$_1$
OR = || ▭ ||$_2$

Classification:

- Distances in Weighted Features Space

$$d\left(\vec{P}_{A'}, \vec{P}_{A_i}\right) = \left[\sum_{k=1}^{N}\left(w_k\,(P_{A'}^{k} - P_{A_i}^{k})\right)^2\right]^{0.5}$$

- Screening of candidates

$$d_2\left(\vec{P}_{A'}, \vec{P}_{A_i}\right) < \varepsilon$$

(A hypersphere in P^N of radius ε)

Classification:

- Distance in Membership Distributions Space

$$d_B\left(f_1(u_k), f_2(u_k)\right) = -\ln[R]$$

$$d_C\left(f_1(u_k), f_2(u_k)\right) = [1 - R]^{0.5}$$

not a metric

where R is the *Bhattacharyya Coefficient*

$$R\left[f_1(u_k), f_2(u_k)\right] = \sum_{k=1}^{D}\left[f_1(u_k) \times f_2(u_k)\right]^{0.5}$$

and

$$f_j(u_k) = \frac{A_j(u_k)}{\sum_{k=1}^{D} A_j(u_k)}$$

for j =1,2

Classification:

- The membership functions have been normalized to have its area equal to one

$$f_j(u_k) = \frac{A_j(u_k)}{\sum_{k=1}^{D} A_j(u_k)} \rightarrow \sum_{k=1}^{D} f_j(u_k) = 1$$

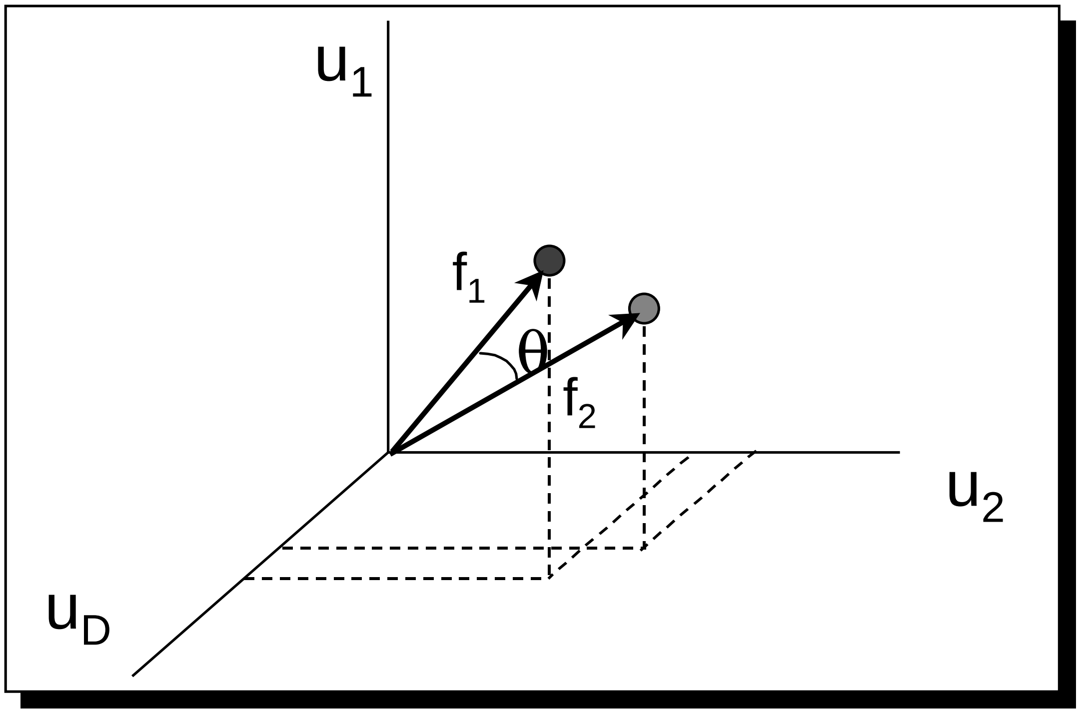
- Geometric Interpretation: Let

$$\sqrt{f_j(u_k)} \quad \text{for } j = 1,2 \text{ and } \forall u_k \in U$$

be the direction cosine on j[th] dimension of two vectors in U (where |U| = D) - The sum of their square equals 1

Classification:

- *Then the Bhattacharyya Coefficient R is the cosine of the angle formed by the two vectors*



when $f_1 = f_2$, $\theta = 0^o$, and $R = \cos\theta = 1$, thus $d_c(f_1, f_2) = 0$

Weights Determination (Saaty)

- Each Weight w_i is the ratio:

$$w_i = \frac{I_i}{Range_i}$$

where:

$$Range_i = (\max_{k=1}^{n} p_k^i) - (\min_{k=1}^{n} p_k^i)$$

$$I_i = \text{Importance of feature } p_i$$

Saaty's method is used to elicit I_i

Reformulation for Parametrized Membership Funct.

- If the membership of the unlabelled fuzzy set A'(u) is defined by the 4-tuple (a', b', α', β')
- and if the semantics of each sentence S_i, in the finite target language $L(G_1)$, is represented by a similar 4-tuple (a_i, b_i, α_i, β_i),
- then we could define:

$$LA[A'] = S_j \in L(G_1) \text{ such that } D(A', A_j) = Min_{i=1}^{n}\{D(A', A_i)\}$$

where:

$$D(A', A_i) = w_1\left(|a' - a_i| + |b' - b_i|\right) + w_2\left(|\alpha' - \alpha_i| + |\beta' - \beta_i|\right)$$

Reformulation for Parametrized Membership Funct. (cont.)

- *Distance*
  - *d* could be another similar metric (e.g. a Weighted Euclidean instead of a Weighted Hamming Distance)
- *Weights*
  - w_1 and w_2 represent the weights for matching the core and support, respectively:
    - w_1 weighs mostly the location of the set (determined by the placement of the core)
    - w_2 weighs mostly the spread of the set (determined by the placement of the support)

LA Applications (1)

1) Infinite Language -> Finite Language

– Approximate an infinite language $L(G_0)$, generated by a grammar G_0 containing recursive non-terminals, with a finite language $L(G_1)$

- The term represented by a sentence in $L(G_0)$ is treated as an unlabelled term. The sentence could be arbitrarily long and therefore difficult to understand
- Design a grammar G_1 generating a finite language $L(G_1)$ that provides sufficient coverage of the element of discourse (descriptors)
- Select the sentence from from $L(G_1)$ with the closest meaning

LA Applications (2)

2) Interpretation of results from AR processes

– Given:

$$R = \bigcup_{i=1}^{n} R_i = \bigcup_{i=1}^{n} (\vec{X}_i \rightarrow Y_i)$$

– We may provide a linguistic answer to the question:

$$\text{If } \left(\vec{X} \text{ is } \vec{X}'\right) \text{ then what is the output } Y'$$

– which is:

$$LA\left[\vec{X}' \bullet R\right]$$

LA Applications (3)

3) Interpretation of Truth Qualification Results

- Given the Possibility Function:

$$(X \text{ is } F) \rightarrow \prod_X = F$$

- Then, its Truth Qualification

$$(X \text{ is } F) \text{ is } \tau \rightarrow \prod_X = F^*$$

  - F and F* are fuzzy sets in U, τ is a fuzzy set in [0,1]
  - F*(u) = τ(F(u)) – using the extension principle
  - The interpretation is: LA[F*(u)]

LA Applications (4)

4) Interpretation of Extension Principle Results

– Given the mapping:

$$* : U \times V \rightarrow W \quad \text{and} \quad A \subseteq U, B \subseteq V$$

– Then the result of applying the mapping * to *A* and *B* can be interpreted as:

$$A * B = LA\left[Sup_{w=u*v} Min(A(u), B(v))\right]$$

LA Applications (5)

5) Interpretation of Results of a Syntactic Substitution Process

- See Riddle example