ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

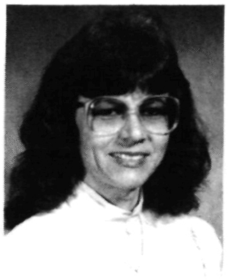
65

5,5-63

15315

P-33

N91-71365



Maria Zemankova, Ph.D.

University of Tennessee
Knoxville, Tennessee

Dr. Zemankova received her B.S. degree from the American University in Cairo in 1977, majoring in mathematics and computer science with a minor in psychology. Her M.S. and Ph.D. degrees were in computer science from Florida State University in 1979 and 1983, respectively. In 1983, she worked on the design and implementation of a relational data base "team-up" at Unlimited Processing, Inc., in Jacksonville, Florida. Since 1984, Dr. Zemankova has been an assistant professor in the Department of Computer Science at the University of Tennessee, Knoxville, and is also engaged in research collaboration with the Oak Ridge National Laboratory. She spent January to August 1987 as a visiting research assistant professor in the Department of Computer Science, University of Illinois, Urbana. Her research interests include intelligent information systems, machine learning, reasoning under uncertainty, and knowledge representation. Dr. Zemankova has authored a book entitled "Fuzzy Relational Data Bases - a Key to Expert Systems" (1984), and several papers on fuzzy intelligent information systems, and on applications of fuzzy set and rough set theories to information systems design and machine learning. Her professional affiliations include AAI, ACM, IEEE, IFSA, NAFIPS, and Sigma xi, and she is an associate editor of the International Journal of Approximate Reasoning.

INTELLIGENT INFORMATION SYSTEMS WITH LEARNING CAPABILITIES

Abstract

An intelligent information system is designed to derive information (that may not be explicitly stored in the data base) by application of rules for inferring plausible answers to queries. The system is divided into the knowledge base (KB) and the inference engine (IE). The KB can be further partitioned into a factual base (FB) and an explanatory base (EB). The FB is used for storing facts (data) that may be imprecise or incomplete, and the EB contains knowledge; i.e., flexible (fuzzy) concepts, relationships, or rules that are used to interpret the available data. The IE is designed to perform flexible reasoning. Clearly, the "intelligence" of the system depends on the knowledge available in the KB and the types of inferences that the IE is capable of performing. An experimental system (APPLAUSE) is discussed, together with demonstration of system function in the knowledge acquisition and querying modes, including its explanatory capabilities.

INTELLIGENT INFORMATION SYSTEMS

WITH LEARNING CAPABILITIES

Maria Zemankova

Department of Computer Science
University of Tennessee
Knoxville, Tennessee

THEME

- Make Machines Behave Intelligently

Marks of intelligence:

- Learning Capability
- Reasoning with Insufficient, Unreliable or Imprecise Data
- Reasoning Under Resource Constraints
- Creativity- Discovery

PLAUSIBLE REASONING

- A Core theory- proposed by Collins and Michalski
- Modifications and Implementation

MODEL

- Hierarchical Organization of Knowledge
- Mechanism to Manipulate Incomplete and Uncertain Knowledge Base
- Domain Independent Inference Mechanism
- Theory is Operationalized with Chemical Periodic Table as a Test Domain

KNOWLEDGE REPRESENTATION

ELEMENTS OF EXPRESSIONS

- Arguments
- Descriptors
- References
- Terms
- Facts - veracity-μ, frequency-ϕ, confidence-γ
- Dependency - forward-backward dependencies
- Implication - forward-backward implications
- Hierarchies - generalization, specialization
- Similarity - context, dominance, typicality

ELEMENTS OF EXPRESSIONS:

Descriptor: d.

breed	attribute
temperature	attribute, function
flies	predicate

Terms: $d_1(a_1)$, or $d_2(a_1, a_2, \ldots)$

$breed(Fido)$
$temperature(latitude,\ altitude)$
$temperature(place)$

References: $r_1, \{r_1, ...\}$.

4	integer
true	logical
*group*6	hierarchical

Factual statements: $d_1(a_1) = r_1 : [\mu, \gamma_\mu, \phi, \gamma_\phi]$

- μ- veracity: Veracity indicates degree with which reference r_1 is applicable to descriptor-argument pair.
- ϕ- frequency: Frequency indicates proportion of argument for which the reference is a valid description of the descriptor-argument pair.
- γ_μ, γ_ϕ- confidences in μ, ϕ.

Examples:

$$density(aluminum) = 2.7 : [1, .99, 1, 1]$$
$$is\_old(john) = yes : [.7, .9, 1, 1]$$
$$engine\_type(car) = 4\_cylinder : [1, 1, .8, .95]$$

Dependency between terms:

$$d_1(a_1) \longleftrightarrow d_2(a_2) : [\alpha, \gamma_\alpha, \beta, \gamma_\beta]$$

- α, β- forward and backward dependency strengths

$$is\_philosopher(X) \longleftrightarrow is\_greek(X) : [.5, .8, .0001, .8]$$

Implications between factual statements:

$$d_1(a_1) = r_1 \Longleftrightarrow d_2(a_2) = r_2 : [\alpha, \gamma_\alpha, \beta, \gamma_\beta]$$

$$grain(place) = rice \Longleftrightarrow$$
$$rain(place) = [80..120in] : [.9, .9, .5, .8]$$

The implications can also be encoded by functions

$$d_1(a_1) = r_1 \Longleftrightarrow d_2(a_2) = f(r_1).$$

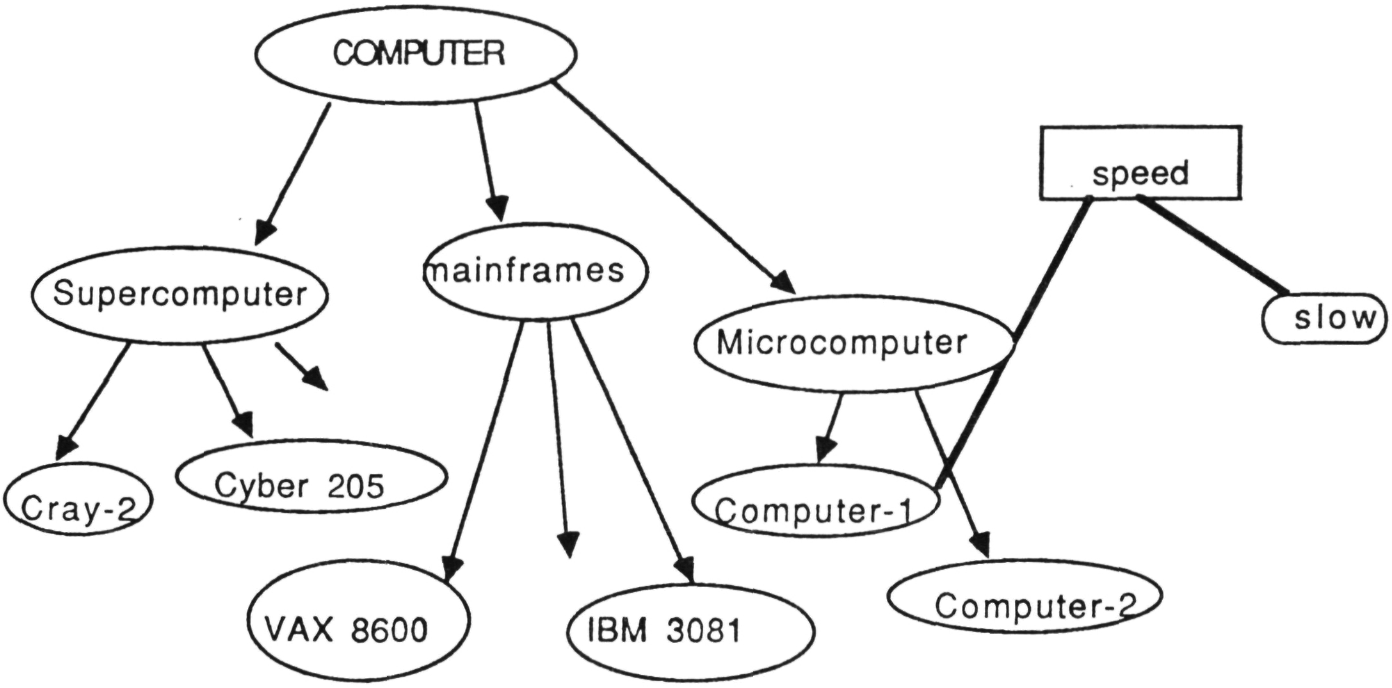
$$radius(circle) = r \Longleftrightarrow$$
$$area(inscribed\ square) = 2r^2 : [1, 1, 1, 1]$$

TRANSFORMATIONS

The transforms A GEN, A SPEC, A SIM, R GEN, R SPEC, R SIM allow traversal in a hierarchy, in the process of inference. Simplified (no parameters) applications of the transforms are given below.

- A GEN

$$
\begin{array}{l}
\text{speed(computer\_1)} = \text{slow} \\
\text{micro\_computer} = \text{gen(computer\_1)}:\ \text{cx} = \text{alu\_size} \\
\text{alu\_size(COMPUTER)} \longleftrightarrow \text{speed(COMPUTER)} \\
\text{micro\_computer} = \text{spec(COMPUTER)} \\
\hline
\text{speed(micro\_computer)} = \text{slow}
\end{array}
$$



A GEN Transformation

- A SPEC

$$\frac{\begin{array}{l}\mathsf{height}(\mathsf{basketball\_player}) = \mathsf{tall} \\ \mathsf{karim} = \mathsf{spec}(\mathsf{basketball\_player})\end{array}}{\mathsf{height}(\mathsf{karim}) = \mathsf{tall}}$$

- A SIM

$$\frac{\begin{array}{l}\text{economy(singapore)} = \text{strong} \\ \text{hongkong} = \text{sim(singapore)}:\ \text{cx} = \text{economic structure}\end{array}}{\text{economy(hongkong)} = \text{strong}}$$

- R GEN

$$\frac{\begin{array}{l}\mathrm{reacts\_with}(\mathrm{potassium}) = \mathrm{chlorine}\\ \mathrm{group7} = \mathrm{gen}(\mathrm{chlorine})\end{array}}{\mathrm{reacts\_with}(\mathrm{potassium}) = \mathrm{group7}}$$

- R SPEC

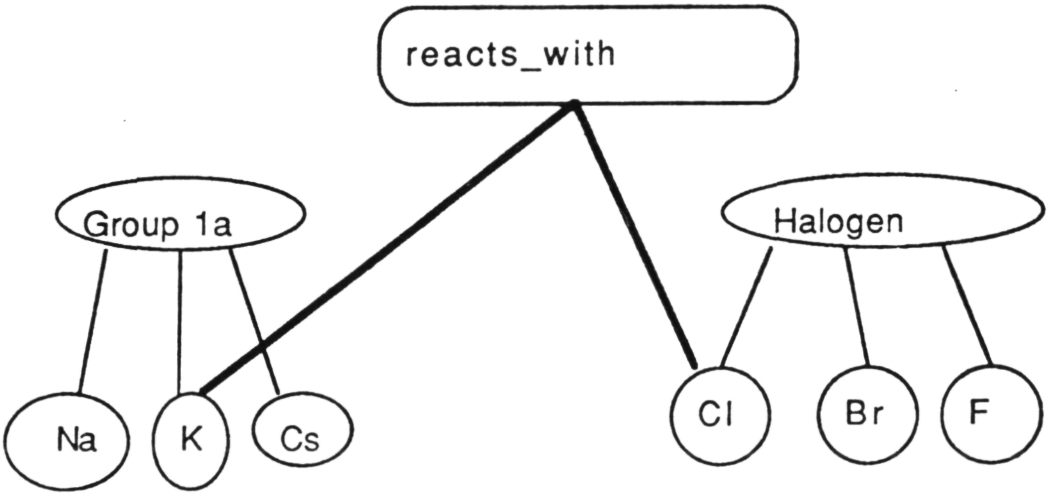
$$\frac{\mathsf{likes(mary)} = \mathsf{carbonated\_drinks} \qquad \mathsf{coke} = \mathsf{spec(carbonated\_drink)}}{\mathsf{likes(mary)} = \mathsf{coke}}$$

- R SIM

$$\frac{\begin{array}{l}\mathsf{habitat}(\mathsf{whales}) = \mathsf{atlantic\_ocean} \\ \mathsf{pacific\_ocean} = \mathsf{sim}(\mathsf{atlantic\_ocean})\end{array}}{\mathsf{habitat}(\mathsf{whales}) = \mathsf{pacific\_ocean}}$$

- Combination of A SIM and R GEN

$$\begin{array}{ll}
\text{reacts_with(potassium)} = \text{chlorine} & \\
\text{sodium} = \text{sim(potassium)} & \text{A SIM} \\
\hline
\text{reacts_with(sodium)} = \text{chlorine} & \\
\text{group7} = \text{gen(chlorine)} & \text{R GEN} \\
\hline
\text{reacts_with(sodium)} = \text{group7} &
\end{array}$$



A SIM, R GEN Combination

Theory of Plausible Reasoning and its Implementation.

Collins and Michalski introduced a theory to model human plausible reasoning. APPLAUSE is an implementation of an extended and modified version of the theory. The methodology is eminently suitable to reason in the domains where knowledge is organized hierarchically. The theory provides mechanisms to manipulate the knowledge base in case of incomplete and uncertain knowledge. Some features of the theory are highlighted with examples from chemical periodic table.

QUERIES

Form:

$$descriptor(argument) \; = \; ref?/[\mu, \gamma_\mu, \phi, \gamma_\phi]? \qquad (1)$$

Aim:

In query (1) the system is to retrieve best reference value together with the estimated parameters. The best reference is one with highest $\mu * \gamma_\mu * \phi * \gamma_\phi$ product.

- Type checking is performed for arguments and descriptors and references when applicable.

ALGORITHM for processing Queries:

- get_query(Q)
- if (get_fact(Q) successful) then – report retrieved information, **exit.**
- elseif reasoning_depth_counter $>$ depth_limit then
  – combine whatever evidence available and exit.
- else
  - increment depth_counter by one.
  - Dep := set of dependencies/implications, such that descriptor occurs in RHS and $\alpha * \gamma_\alpha >$ threshold T.
  - sort dependencies and implications according to decreasing $\alpha * \gamma_\alpha$ (gather strongest evidence first).
  - repeat until no more dependencies.
    * evaluate LHS of dependency or implication. If necessary call this routine to evaluate LHS.
    * apply suitable transforms such as A GEN, A SPEC, A SIM and compute RHS, decrement depth_counter by one and exit.
- combine evidences:
  - choose best γ_μ, γ_ϕ or $\mu * \gamma_\mu, \phi * \gamma_\phi$ products, for type 1 or type 2 query respectively.
  - compute union and intersection of ranges to give upper and lower bounds on the range of the conclusion respectively.

EXAMPLE

Given:

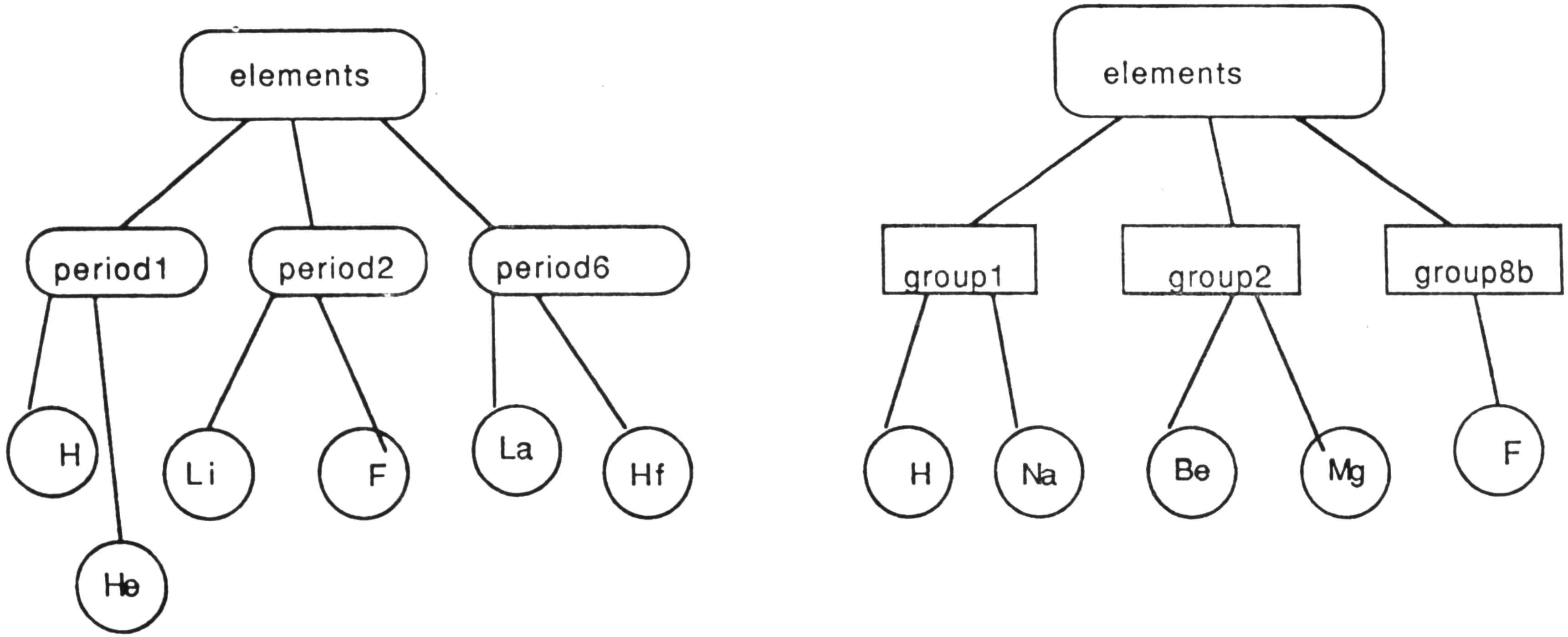
- Group8a consists of gases [He, Ne, Ar, Kr, Xe, Rn].
- Boiling points of only 4 gases are known.

  [He/-269, Ne/-246, Ar/-185, Xe/-108].

Query: Find boiling point of Kr.

Process:

- Statistical analysis will be made to see if it is reasonable to aggregate the boiling points into a range and propagate it to a parent node.
- Kr has two parents, group8a and period4.
- Suitability of generalization is tested in both hierarchies.
- The better one is selected for inference.



Two possible hierarchies for periodic table

Criteria for generalization:

- Low standard deviation of residuals
- Generalize from large number of points.
- Low range of residual errors (fewer outliers)
- Presence of functional dependencies with characteristics similar to those in the neighboring classes.
- 'Causal connection'

Criteria for generalization:

- Low standard deviation of residuals
- Generalize from large number of points.
- Low range of residual errors (fewer outliers)
- Presence of functional dependencies with characteristics similar to those in the neighboring classes.
- 'Causal connection'

Total # of Points = 100
Total # of Variables = 3
Names of the attributes GROUP PERIOD ATNUM
A suitable hierarchy is to be decided for attribute ATNUM

Evaluating Group as the Primary Attribute for Classification
Total Number of Distinct Classes = 18

Class #	Group #	Group	Attr Max	Attr Min	Attr Avg	Std dev of residuals as % of AttrAvg
1	1	IA	87	1	30.43	35.9
2	2	IIA	88	4	36.33	24.6
3	2.1	IIIB	100	21	74.27	6.3
4	2.2	IVB	72	22	44.66	12.8
5	2.3	VB	73	23	45.66	12.2
6	2.4	VIB	74	24	46.66	12.2
7	2.5	VIIB	75	25	47.66	12.0
8	2.6	VIII	76	26	48.66	11.7
9	2.65	VIII	77	27	49.66	11.5
10	2.7	VIII	78	28	50.66	11.3
11	2.8	IB	79	29	51.66	11.0
12	2.9	IIB	80	30	52.66	10.8
13	3	IIIA	81	5	35.8	21.3
14	4	IVA	82	6	36.8	20.7
15	5	VA	83	7	37.8	20.2
16	6	VIA	84	8	38.8	19.6
17	7	VIIA	85	9	39.8	19.1
18	8	VIIIA	86	2	34.33	26.1

Evaluating Period as the Primary Attribute for Classification
Total Number of Distinct Classes = 7

Class#	Period#	AttrMax	AttrMin	AttrAvg	Std dev of residuals as % of AttrAvg
1	1	2	1	1.5	26.1
2	2	10	3	6.5	0.0
3	3	18	11	14.5	0.0
4	4	36	19	27.5	2.7
5	5	54	37	45.5	2.7
6	6	86	55	70.5	6.5
7	7	100	87	93.5	3.8

Generalization	group8a	period4
BP range	[-108 .. -269]	[58 .. 3450]
slope m	42	-416
std dev of residuals σ_r	5%	46%
% intersection with neighboring class x	60%	100%
number of points n	4	17
Computed α	0.88	0.8
Computed γ_α	0.93	.3

The equation (implication in a functional form) derived by best line fit method:

BP = -317 + 41.8*period (valid for group8a).

The parameters α and γ_α are estimated by evaluating compliance to the criteria for generalizing.

$$\alpha = (1 - 0.2 * x)$$

$$\gamma_\alpha = 0.5 * \left(\frac{m_{max} - m}{m_{max} - m_{min}}\right) + 0.4 * (1 - \sigma_r) + 0.1 * \left(\frac{n}{n+4}\right)$$

Tabulated factors favor generalization in group8a rather than in period 4.

PARAMETERS FOR THE DERIVED CONCLUSION:

Derivation using A SPEC without functional dependency

$$BP(Kr) = [-108, -269]$$

Parameters $[\mu, \gamma_\mu, \phi, \gamma_\phi]$ are directly inherited from the parent node, however the precision of the answer is low.

The answer is made more precise (narrower range) by using functional dependencies discovered in the related elements.

Derivation using A SPEC together with functional dependency.

Assume τ and γ_τ for Kr within group8a = .9, .95

These can be estimated by evaluating common relevant features among the siblings.

BP(Kr) = -317 + 41.8*4 = -149.8

$$\begin{aligned}
\mu_c &= \mu_1 = 1 \\
\gamma_{\mu_c} &= \gamma_\mu * \alpha * \gamma_\alpha * \tau * \gamma_\tau \\
&= 1 * .9 * .9 * .9 * .95 = 0.6925 \\
\phi_c &= \phi_1 = 1 \\
\gamma_{\phi_c} &= \gamma_{\phi_1} * \alpha * \gamma_\alpha * \tau * \gamma_\tau \\
&= 1 * .9 * .9 * .9 * .95 = 0.6925
\end{aligned}$$

Derivation using A SIM transform.

Find elements similar to Kr in some context which affects boiling point.

Suppose, relevant context is

CX = (.7*group + .3*period) Rule 1

and the dependency is given by,

CX -> boiling point: $\alpha = .75, \gamma_\alpha = 1$; Rule 2

- α and γ_α estimated by global multiple regression analysis.
- Localize the search space within the neighborhood of the argument in question
- Similarity σ and γ_σ are computed according to the formulas:

$$\sigma(Arg_1, Arg_2) = \sum W_i * \sigma(attr_i(Arg_1), attr_i(Arg_2))$$

$$\gamma_\sigma(Arg_1, Arg_2) = \sum W_i * \gamma_\sigma(attr_i(Arg_1), attr_i(Arg_2))$$

where, the weights W_i are normalized such that the sum of weights is 1.

Assuming pairwise similarity σ and γ_σ values

$$\sigma(\text{gr8a, gr7a}) = .2; \;\; \gamma_\sigma = .95$$
$$\sigma(\text{per4, per3}) = .8; \;\; \gamma_\sigma = .95$$
$$\sigma(\text{per4, per5}) = .7; \;\; \gamma_\sigma = .95$$

and W_i given by context in Rule 1, we get

Elem	Gr.	Per.	σ(Kr, Elem)	γ_σ(Kr, Elem)
Ar	8a	3	$.7*1 + .3*.8 = .94$	.95
Xe	8a	5	$.7*1 + .3*.7 = .91$	.95
Cl	7a	3	$.7*.2 + .3*.8 = .38$	.95
Br	7a	4	$.7*.2 + .3*1 = .44$	.95
I	7a	5	$.7*.2 + .3*.7 = .35$	.95

Disregard elements with $\sigma * \gamma_\sigma <$ threshold T.

Similarity transform reference and parameter estimation:

$$\text{BP(Kr)} = \text{BP(Element)}\ [\mu, \gamma_\mu, \phi, \gamma_\phi]$$

Ar: BP(Ar) = -185 $[\mu = 1, \gamma_\mu = 1, \phi = 1, \gamma_\phi = 1]$

$$\begin{aligned}
&\text{BP(Kr)} = \text{BP(Ar)} = \text{-185.94} \\
&\mu_c = \mu = 1, \\
&\gamma_{\mu_c} = \gamma_\mu * \sigma * \gamma_\sigma * \alpha * \gamma_\alpha \\
&\quad = 1 * .94 * .95 * .7 * 1 = .625 \\
&\phi_c = \phi = 1 \\
&\gamma_{\phi_c} = \gamma_\phi * \sigma * \gamma_\sigma * \alpha * \gamma_\alpha
\end{aligned}$$

$$1 * .94 * .95 * .7 * 1 = .625$$

similarly,

Xe: BP(Xe) = -108 [1, 1, 1, 1]

$$\begin{aligned}
&\text{BP(Kr)} = \text{BP(Xe)} = \text{-108.91} \\
&\mu = 1 \\
&\gamma_\mu = 1 * .91 * .95 * .7 * 1 = .605 \\
&\phi_c = 1 \\
&\gamma_{\phi_c} = 1 * .91 * .95 * .7 * 1 = .605
\end{aligned}$$

COMBINATION OF EVIDENCES

- Take the reference value of the result as the weighted average value of the BPs where the weights are decided by $\sigma * \gamma_\sigma$ product.
- The parameters are taken as the weighted average of the evidences.

BP(Kr)= [(BP(Ar)$*\sigma$(Kr,Ar) $+$ BP(Xe)$*\sigma$(Kr,Xe)]/$(\Sigma\,\sigma_i)$

BP(Kr) $= (-185 * .94 + -108 * .91)/(.94 + .91) = -147.1$

$\mu_c = \Sigma\,\mu_i/N$ (if $\Delta\mu$ not large)

$(1+1)/2 = 1$

$$\gamma_{\mu_c} = \Sigma\,\gamma_{\mu_i} * \sigma * \alpha * \gamma_\alpha/N$$

$$= (1 * .94 * .75 * 1 + 1 * .91 * .75 * 1)/(1+1) = .6175$$

$\phi_c = \Sigma\,\phi_i/N$

$(1+1)/2 = 1$

$$\gamma_{\phi_c} = \Sigma\,\gamma_{\phi_i} * \sigma * \alpha * \gamma_\alpha/N$$

$$= (1 * .94 * .75 * 1 + 1 * .91 * .75 * 1) = .6175$$

Comparison of parameters:

Parameter	A SPEC	A SIM	Actual
BP	-149.8	-147	-152
μ	1	1	-
γ_μ	.6925	.6175	-
ϕ	1	1	-
γ_ϕ	.6925	.6175	-

Choose results obtained by A SPEC since it yields inference with higher confidence.

CONCLUSION

Plausible Reasoning provides a useful mechanism to manipulate available knowledge base to infer conclusions not arrivable by traditional logic.