Weather Prediction Using Case-Based Reasoning and Fuzzy Set Theory

Bjarne Hansen

Master of Computer Science Student
DalTech, Dalhousie University

Meteorologist
Meteorological Service of Canada

International Conference on Case-Based Reasoning
Vancouver, BC, Canada – 31, July, 2001

Outline

Introduction, Literature Review and Synthesis

- ***Case-Based Reasoning***
- ***Fuzzy k-Nearest Neighbor Technique***
- ***Airport Weather Prediction Problem***
- ***Hypothesis***

System for Case-Based Weather Prediction

Five Sets of Experiments

Conclusions

Case-Based Reasoning

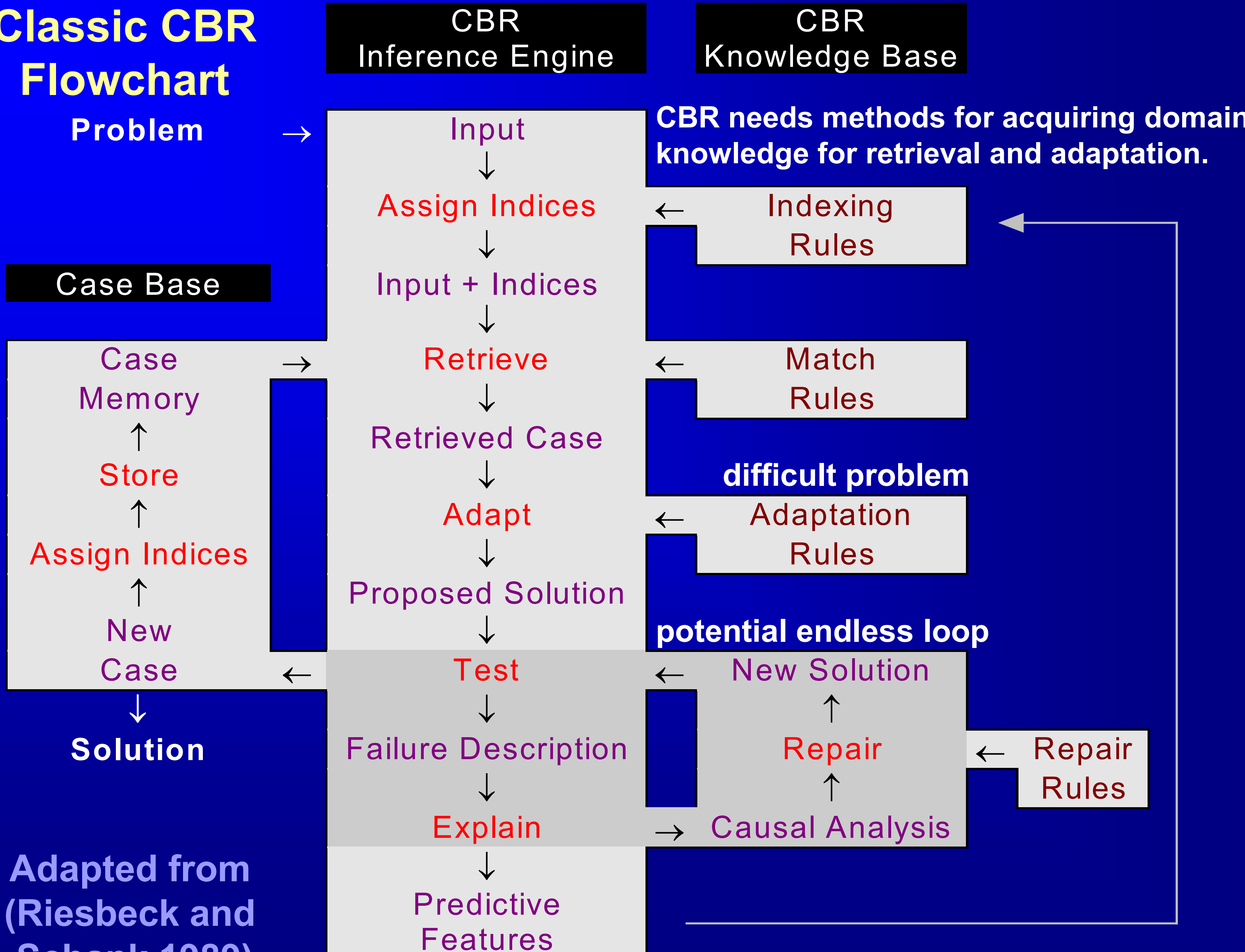
"A case-based reasoner solves new problems by adapting solutions that were used to solve old problems." (Riesbeck and Schank 1989)

CBR = retrieval + analogy + adaptation + learning (Leake 1996)

CBR is very effective in situations "where the acquisition of the case-base and the determination of the features is straightforward compared with the task of developing the reasoning mechanism." (Cunningham and Bonzano 1999)

Requirement for application-specific knowledge to handcraft cases creates a bottleneck in CBR development. Domain experts are too expensive to employ for construction and maintenance of decision support systems. (Aha 1997)

Classic CBR Flowchart



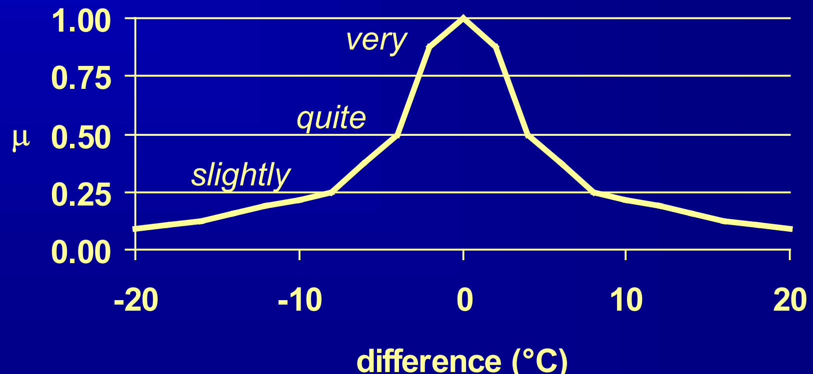
Adapted from (Riesbeck and Schank 1989)

Fuzzy Logic *

Fuzzy logic provides a theoretically sound set of techniques that have been used for computational modeling in many domains.

- Fuzzy logic is effective for eliciting and encoding knowledge that can control recognition of similarity between two weather situations. (Hansen 1997)
- Fuzzy logic is often used to model continuous, real-world systems. Many systems dealing with environmental data use fuzzy logic. (Hansen et al. 1999)

For example: fuzzy set to describe *degree of similarity* of temperatures



* Definition: "Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth – truth values between 'completely true' and 'completely false.'" (*Fuzzy Logic FAQ* 1993)

k-Nearest Neighbors

- Definition: for a particular point in question, in a population of points, the *k* nearest points. (Dudani 1976)

  "It is reasonable to assume that observations which are close together (according to some appropriate metric) will have the same classification. Furthermore, it is also reasonable to say that one might wish to weight the evidence of a neighbor close to an unclassified observation more heavily than the weight of another neighbor which is at a greater distance from the unclassified observation." (Dudani 1976)

- Basic CBR method. Commonly used to explain an observation when there is no other more effective method. (Aha 1998)

Fuzzy *k*-Nearest Neighbor Technique

- Nearest neighbor technique in which the basic measurement technique is fuzzy. (Keller 1985)

Two Improvements to *k*-nn Technique Achieved by Using Fuzzy *k*-nn Approach

- "Improve performance of retrieval in terms of accuracy because of avoidance of unrealistic absolute classification." (Keller 1985)
- "Increase the interpretability of results of retrieval because the overall degree of membership of a case in a class that provides a level of assurance to accompany the classification." (Keller 1985)

Current Research in CBR

Scaling Up

The greatest opportunity for the development of CBR systems is scaling up systems and integrating them with existing large databases. (Kamp et al. 1998)

Integrating Domain Knowledge to Aid Retrieval

"A topic of future research within intelligent retrieval is the integration of domain knowledge and background knowledge to enhance the semantic of the retrieval. This could be done by considering and integrating techniques from knowledge representation [and] in this area, further research includes finding guidelines for finding the right tradeoff between expressiveness and complexity for different application scenarios, the search for approximations..." (Kamp et al. 1998)

Current Research in CBR (contd.)

Integrating Domain Knowledge to Aid Adaptation

"Central questions for adaptation are which aspects of a situation to adapt, which changes are reasonable for adapting them, and how to control the adaptation process. Answering these questions may require considerable domain knowledge, which in turn raises the questions of how to acquire that knowledge. Many CBR systems depend on that knowledge being encoded *a priori* into rule-based production systems. Unfortunately, this approach raises the same types of knowledge acquisition issues that CBR was aimed at avoiding. It has proven a serious impediment to automatic adaptation." (Leake 1996)

Weather Prediction

There are two methods to predict weather:

- ***Dynamical approach*** **- based upon equations of the atmosphere, uses finite element techniques, and is commonly referred to as "computer modeling" or "numerical weather prediction."**
- ***Empirical approach*** **- based upon the occurrence of analogs, or similar weather situations.**

(Lorenz 1969)

Analog Forecasting – Case-Based Prediction

- **Depends on retrieval of similar cases.**
- **Method: make prediction for the present case based on the outcome of similar past cases.**

(*Online Guide to Forecasting* 2000)

Airport Weather Prediction – Definition

- A concise statement of the expected weather conditions at an airport during a specified period. (US National Weather Service 1999)
- Commonly referred to as TAF – *Terminal Aerodrome Forecast*. When pilots give weather forecasts to passengers before landing, they are reading TAFs. (US National Weather Service 1999)
- TAFs are the most precise and the most challenging type of forecast to make: height of low cloud ceiling should be accurate to within 100 feet; horizontal visibility on ground, when low, should be accurate to within 400 metres; time of change from one flying category to another expected to be accurate to within one hour. (*MANAIR*, Meteorological Service of Canada, 1998)

Importance of Airport Weather Predictions

Safety

- Pilots load on extra fuel when low cloud and poor visibility are forecast at their destination, in case they must divert to an alternate airport to land. (Patton 1996)

Economics

- "The economic benefit of a uniform, hypothetical increase in TAF accuracy of 1% is approximately $1.2 million [Australian] per year for Qantas international flights into Sydney." (Leigh 1995)
- In Canada, production of TAF's accounts for about $5,000,000 per year in revenue to Met. Service from airlines. (Macdonald 1999)

State-of-the-Art of Airport Weather Prediction – Persistence Climatology

> "Persistence climatology (PC) is widely recognized as a formidable benchmark for very short range prediction of ceiling and visibility [critical attributes of airport weather]."
> (Vislocky and Fritsch 1997)

Basic objective of PC is to answer the question: In similar past situations, what were the outcomes 1, 2, 3, ... hours later? PC is a meteorological application of joint probability.

For example:

Suppose that it is 6 am in June and the airport is "socked in" in fog. The flying category is the lowest possible, Category 1. Using PC, tabulate before-the-fact probabilities (prior probabilities) to forecast for such a situation. The database is searched for all instances of {*July*, *6 am*, *flying category 1*}, the flying categories of the subsequent hours are tabulated, and probabilities were prepared accordingly. (Martin 1972)

Limitation in Current Airport Weather Forecasting Systems – Underlying Assumption that Weather is Adequately Described by Categories

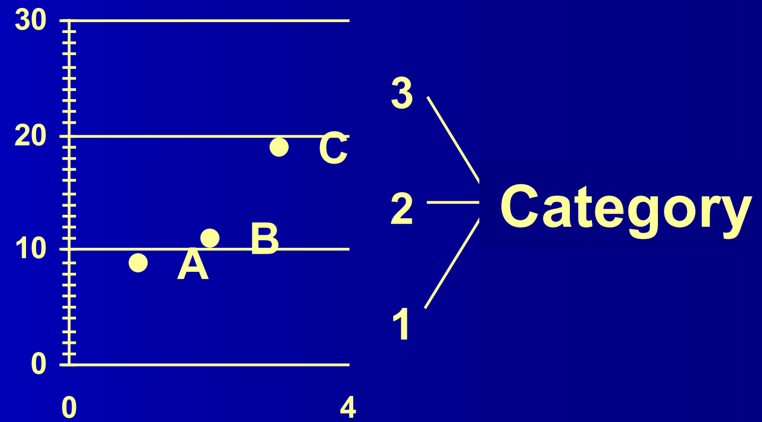
All current systems, both analog based and rule based, are based on the assumption that airport weather data can only be represented and processed indirectly according to categories. Current systems use:

- Category based treatment of variables.
- Prior probability based treatment of situations.

(Clarke 1995, Garner 1995, Gollvik and Olsson 1993, Keller et al. 1995, Kilpinen 1993, Kumar et al. 1994, Meyer 1995, Porter and Seaman 1995, Shakina et al. 1993, Warner and Stoelinga 1995, Vislocky and Fritsch 1997, Whiffen 1993, Wilson and Sarrazin 1989)

Category based treatment of variables

- Does not intuitively reflect degree of similarity between cases.
- Effort to compensate by using finer categories may result in yield “no past event” upon which to base a prediction.



Prior probability based treatment of situations

- Limits the specificity of the situation description. Not practical to calculate prior probabilities of outcomes of a specific situation, such as:

  July 10th, 6 am, ceiling height 100 feet, wind southerly 5 km/h, wind shift three hours hence to westerly 15 km/h

- There are too many possible combinations to account for before the actual event.

Fuzzy *k*-nn Based Airport Weather Prediction – Our Proposal

Unique application for analog weather prediction.

Takes the general query that is the basis of all previous systems:

> Pre-compile probabilities of future weather categories based on outcomes of pre-selected categories of past weather cases, assuming that the pre-selected categories will closely resemble actual future weather cases.

Replaces with a more specific, better targeted database query:

> At "run-time" compile probabilities of future weather values based on the outcomes of specific past cases most similar to the specific present case, and weight each similar past cases according to its degree of similarity with the present case.

Hypothesis

Querying a large database of weather observations for past weather cases similar to a present case using a fuzzy *k*-nearest neighbor algorithm that is designed and tuned with the help of a weather forecasting expert can increase the accuracy of predictions of cloud ceiling and visibility at an airport.

WIND-1 System for for CBR Weather Prediction

WIND-1 **– Weather Is Not Discrete - Version 1.**

Consists of two parts:

- ***Large case base of weather observations*. A weather archive of over 300,000 consecutive hourly weather observations.**
- ***Fuzzy k-nn algorithm.* Measures similarity between temporal cases, past and present intervals of weather observations, and makes forecasts for a present case based on the outcomes of the most similar past cases. Algorithm is tuned with the help of a domain expert, a weather forecaster experienced in noting similarities between cases.**

Large Case Base of Weather Observations

Category	Attribute	Units
temporal	date	Julian date of year (wraps around)
	hour	hours offset from sunrise/sunset
cloud ceiling and visibility	cloud amount(s)	tenths of cloud cover (for each layer)
	cloud ceiling height	height in metres of $\geq$ 6/10ths cloud cover
	visibility	horizontal visibility in metres
wind	wind direction	degrees from true north
	wind speed	knots
precipitation	precipitation type	nil, rain, snow, etc.
	precipitation intensity	nil, light, moderate, heavy
spread and temperature	dew point temperature	degrees Celsius
	dry bulb temperature	degrees Celsius
pressure	pressure trend	kiloPascal $\times$ hour $^{-1}$

Large Case Base of Weather Observations (contd.)

Over 300,000 consecutive hourly obs for Halifax Airport

YY/MM/DD/HH	Ceiling 30's m	Vis km	Wind Directn 10's deg	Wind Speed km/hr	Dry Bulb deg C	Wet Bulb deg C	Dew Point deg C	MSL Press kPa	Station Press kPa	Cloud Amount tenths	WEATHER
64/ 1/ 2/ 0	15	24.1	14	16	-4.4	-4.4	-5.6	101.07	99.31	10	
64/ 1/ 2/ 1	13	6.1	14	26	-2.2	-2.2	-2.8	100.72	98.96	10	ZR-
64/ 1/ 2/ 2	2	8.0	11	26	-1.1	-1.7	-2.2	100.39	98.66	10	ZR-F
64/ 1/ 2/ 3	2	6.4	11	24	0.0	0.0	-0.6	100.09	98.36	10	ZR-F
64/ 1/ 2/ 4	2	4.8	11	32	1.1	1.1	0.6	99.63	97.90	10	R-F
64/ 1/ 2/ 5	2	3.2	14	48	2.8	2.8	2.2	99.20	97.50	10	R-F
64/ 1/ 2/ 6	3	1.2	16	40	3.9	3.9	3.9	98.92	97.22	10	R-F
64/ 1/ 2/ 7	2	2.0	20	40	4.4	4.4	4.4	98.78	97.08	10	F
64/ 1/ 2/ 8	2	4.8	20	35	3.9	3.9	3.3	98.70	97.01	10	F
64/ 1/ 2/ 9	4	4.0	20	29	3.3	3.3	2.8	98.65	96.96	10	R-F
64/ 1/ 2/10	6	8.0	20	35	2.8	2.8	2.2	98.60	96.91	10	F
64/ 1/ 2/11	8	8.0	20	32	2.8	2.2	2.2	98.45	96.77	10	F
64/ 1/ 2/12	9	9.7	23	29	2.2	2.2	1.7	98.43	96.75	10	F
64/ 1/ 2/13	9	11.3	23	32	1.7	1.7	1.1	98.37	96.69	10	

...

6 Megabytes, quality controlled, ready-to-use.

Fuzzy *k*-nn Algorithm

Three steps to construct and use algorithm.

1. **Configure similarity-measuring function.**
2. **Traverse case base to find *k*-nn.**
3. **Make prediction using weighted median of *k*-nn.**

Expertly Configured Similarity-Measuring Function

Expert weather forecaster uses a fuzzy vocabulary to provide knowledge about how to perform case comparisons. Specifies attributes to compare and the order in which they are to be compared. Expert fills in a questionnaire:

Attributes to compare in the order that they should be compared – most discriminating attributes first.

date of the year, hour of the day, cloud amount, cloud ceiling height, visibility, wind direction, wind speed, precipitation type, precipitation intensity, dew point temperature, dry bulb temperature, pressure trend

Expertly Configured Similarity-Measuring Function (contd.)

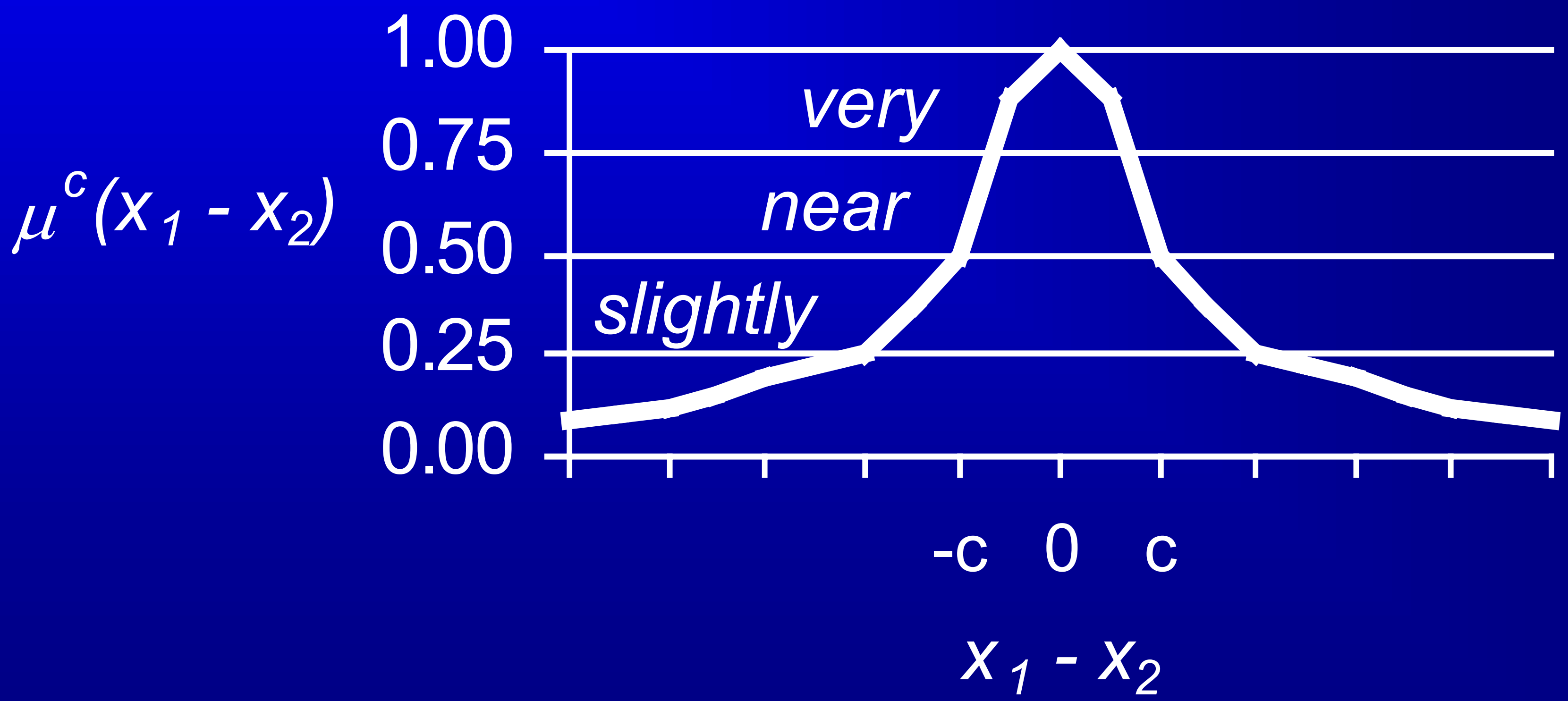
Expert specifies thresholds for various degrees of *near*

Attribute	*slightly near*	*near*	*very near*
date of the year	*60 days*	*30 days*	*10 days*
hour of the day	*2 hours*	*1 hours*	*0.5 hours*
wind direction	*40 degrees*	*20 degrees*	*10 degrees*
dew point temperature	*4 degrees*	*2 degrees*	*1 degree*
dry bulb temperature	*8 degrees*	*4 degrees*	*2 degree*
pressure trend	$0.20\ kPa \cdot hr^{-1}$	$0.10\ kPa \cdot hr^{-1}$	$0.05\ kPa \cdot hr^{-1}$

Comparing Continuous-Number Attributes

Similarity-measuring fuzzy sets made according to specifications.

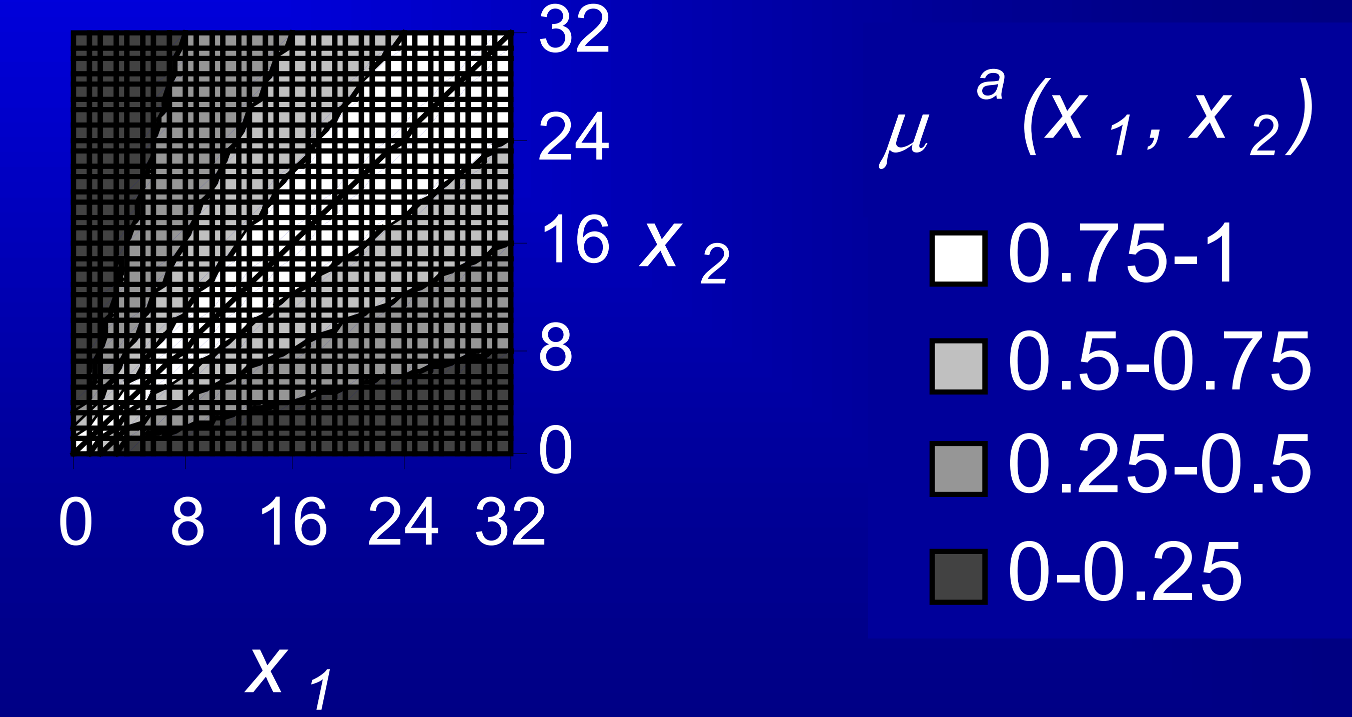
Fuzzy set for describing degrees of similarity between x_1 and x_2.



Comparing "Absolute Number" Attributes

If attributes are limited to the zero-or-above range (e.g., wind speed) then it is their relative magnitudes that are important for matching. They are compared using a modified ratio operation, with special routines to handle for values near zero.

Fuzzy decision surface



Relationships Between Nominal Attributes

For example, for precipitation types:

Fuzzy Relationships

$\mu(\text{type}_1, \text{type}_2)$

	Nil	Drizzle	Showers	Rain
Nil	1.00			
Drizzle	0.02	1.00		
Showers	0.03	0.50	1.00	
Rain	0.01	0.50	0.75	1.00
…	…	…	…	…

Traversing the Case Base to Find *k*-nn

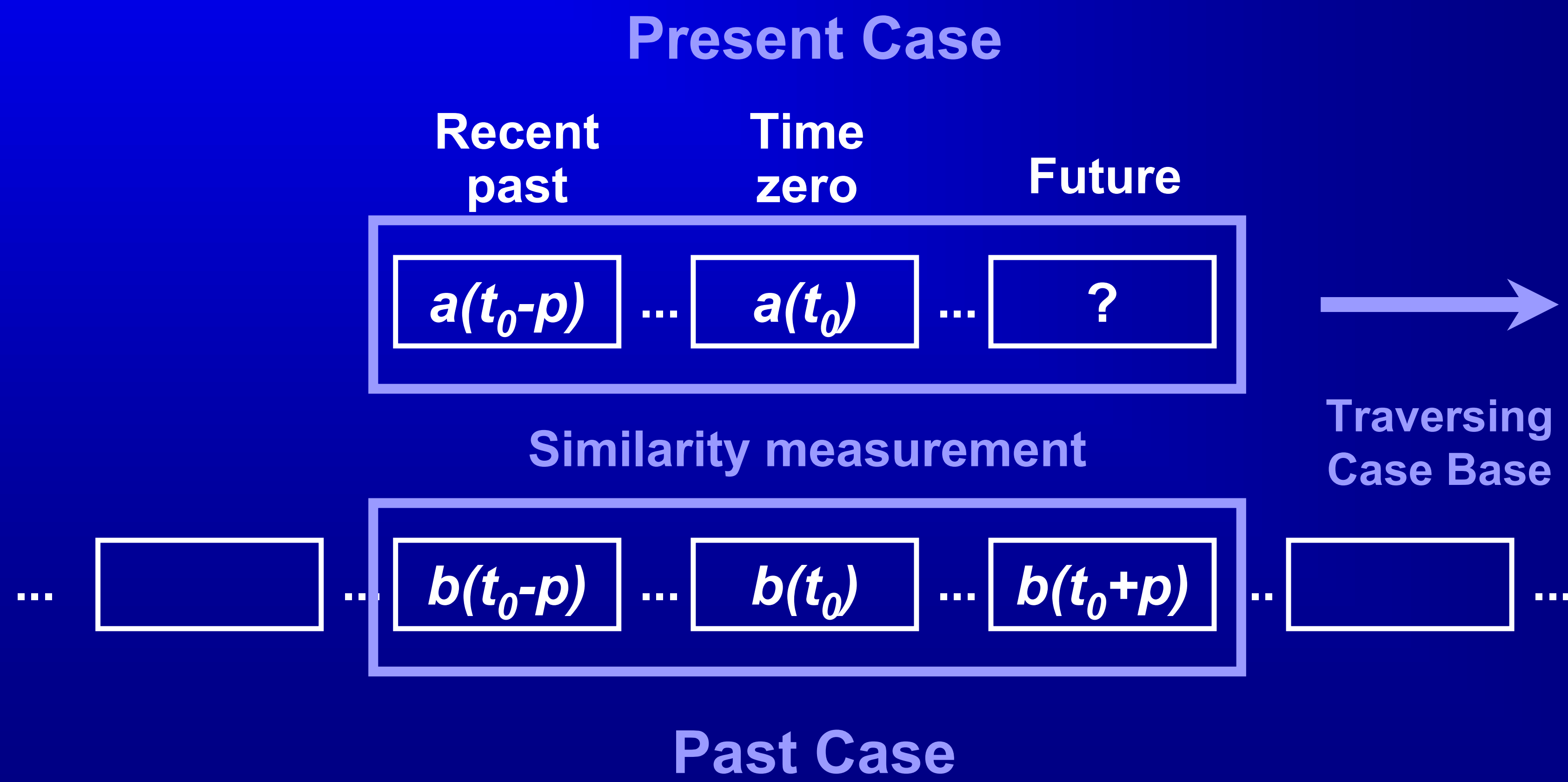
Given a present incomplete case for forecast for
and a case base of past cases to base forecasts upon,
compare present case to past cases one-by-one.

For each hour of two cases being compared, the overall
degree of similarity of their attributes is computed as the
minimum value of for the compared attributes.

Traversing the Case Base to Find *k*-nn (contd.)

Compare present case with past cases

case-to case, hour-to-hour, attribute-to-attribute



Traversing the Case Base to Find *k*-nn (contd.)

Rate past cases according to their overall similarity with present case.

Threshold for admission to the *k*-nn set is α-*level*,
lowest level of similarity among the *k*-nn

$0.0 \leq \alpha \leq 1.0$

α-*level* initialized to 0.0

α-*level* rises during traversal

computational cost of similarity measurement decreases steadily

$O(n^3) \rightarrow O(n)$

In essence, $(1.0 - \alpha)$ is the radius of a contracting hypersphere, centered on the many, expertly described dimensions of the present case, which contains *k*-nn after case base traversal.

Traversing the Case Base to Find *k*-nn (contd.)

Algorithm

α = 0.0

next case

for every past case in the case base

min_similarity = 1.0

for every hour in each case

for every attribute in each hour

x = *sim* (past case, present case)

if *x* < α

stop similarity measurement

min_similarity = *min*(*min_similarity*, *x*)

if *min_similarity* > α

α = *min_similarity*

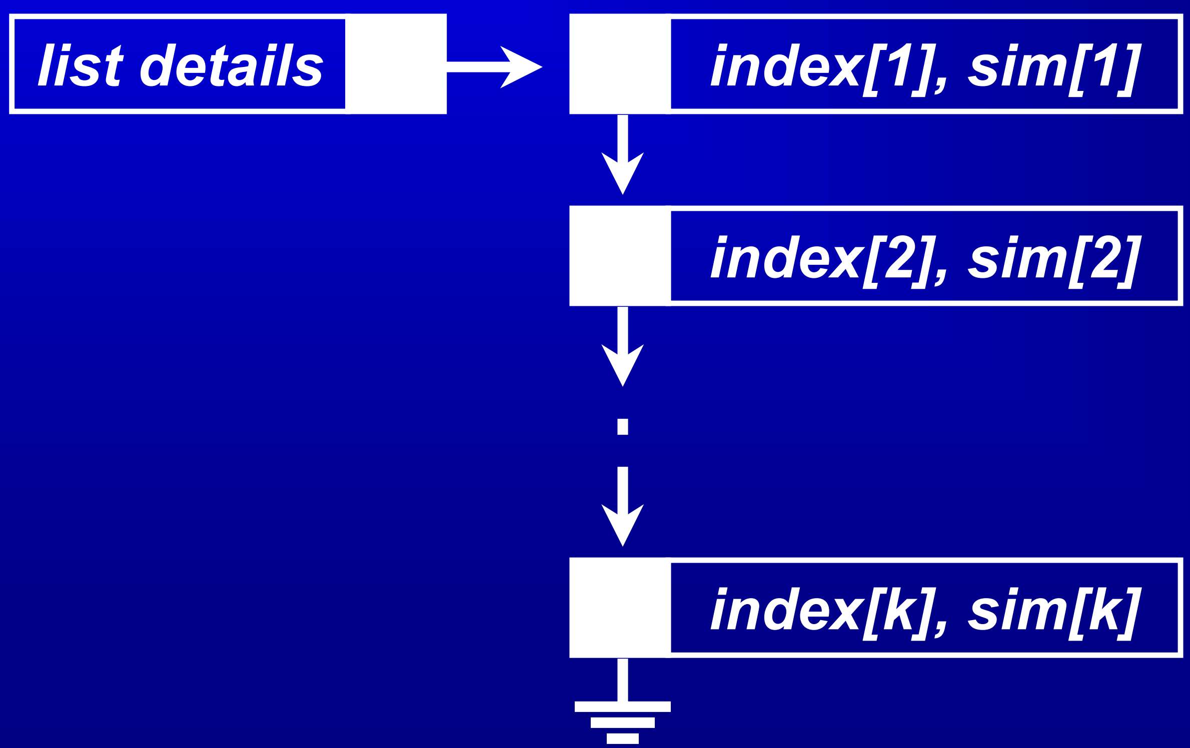
save past case in *k*-nn set

linked list

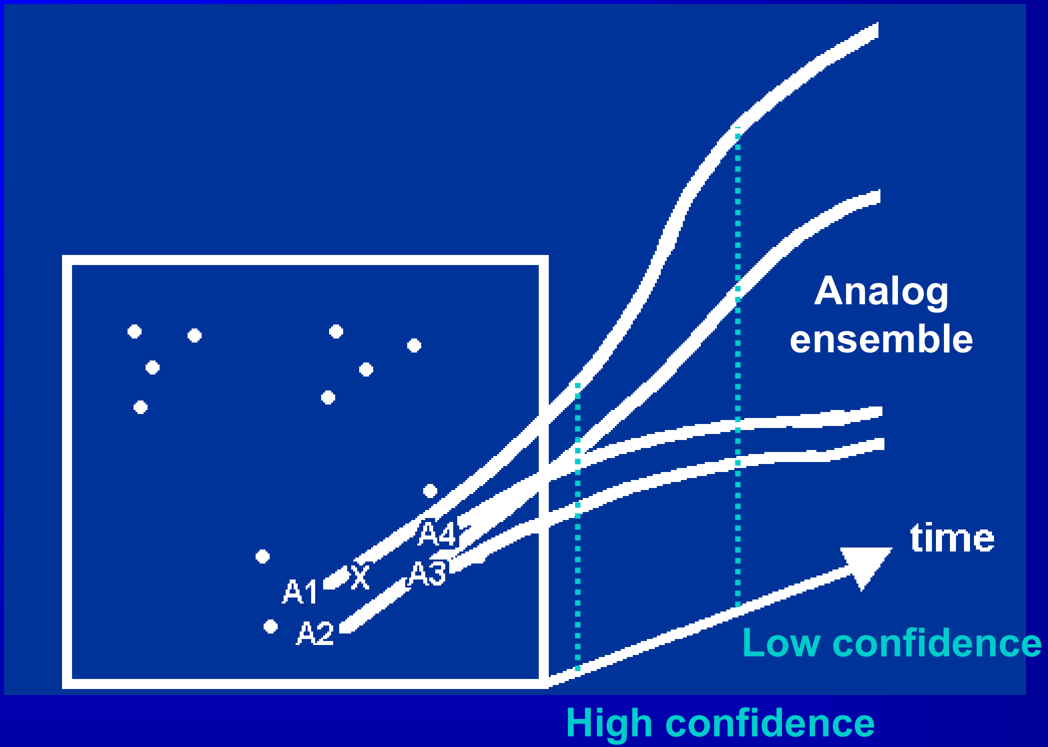
Traversing the Case Base to Find *k*-nn (contd.)

Save most similar past cases in linked list ordered according to degree of similarity.

Threshold for admission = α-*level* = *sim[k]*



Make Prediction Using Weighted Median of *k*-nn



Prediction for new case, X, based on most similar past cases, *A1… A4*.

Each past case weighted according to its level of similarity with *X*.

Confidence determined by intactness or divergence of flow of *A1… A4*.

Experiments

Five sets of realistic forecast simulations.

Used "leave one out method."

- Each experiment tested the effect of varying a system component.
- In most experiments, the first 35 years of weather data, 1961-1995, was used as the case base and the final year of data, 1996, was used as a source of "new cases."
- In each set of experiments, 1000 simulated forecasts were produced.
- For comparison and control, the same 1000 randomly-chosen hours were used in each set of experiments.
- In each simulated forecast, a case is taken from 1996 data and was used as a present case. It is input to *WIND-1*. During the forecast process, the outcome of present case (future part) was hidden.
- WIND-1 produced a forecast for the present case based on the *k*-nn.
- Accuracy of the forecast is verified by comparing the forecast with the actual outcome of the present case.

Verification Method

Each forecast verified using standard measures (Stanski et al. 1999) according to the accuracy of forecasts of three significant flying categories:

Ceiling (m)		Visibility (km)		Flying category
< 200	or	< 3.2	$\Rightarrow$	below alternate
$\geq$ 200	and	$\geq$ 3.2	$\Rightarrow$	alternate
$\geq$ 330	and	$\geq$ 4.8	$\Rightarrow$	VFR

Three sorts of prediction-versus-actual outcomes were counted:

		OBSERVED YES	OBSERVED NO
FORECAST	YES	*hit*	*false alarm*
	NO	*miss*	*non-event*

Verification Method (contd.)

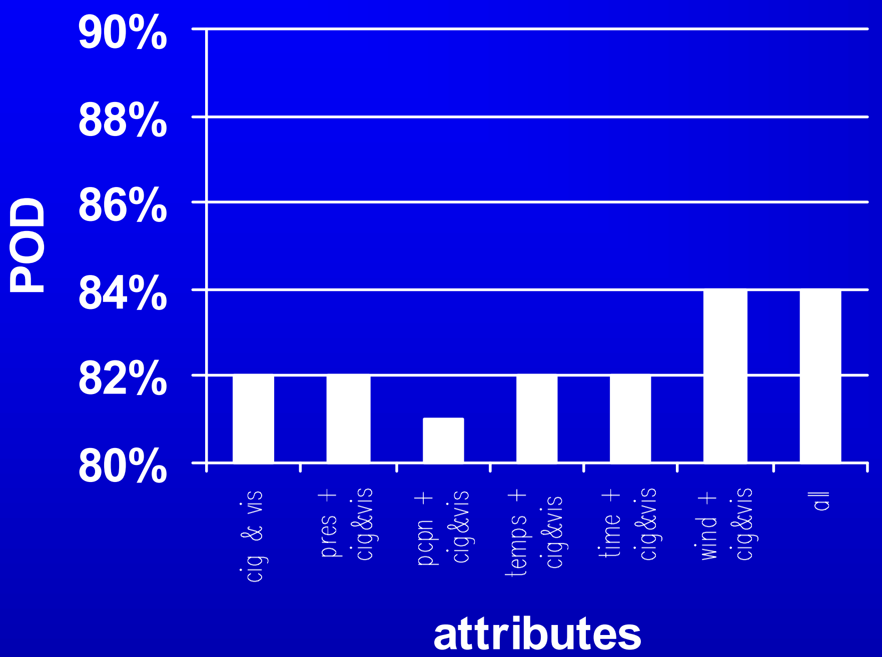
Two statistics are calculated based on cumulative frequencies of hits, misses, and false alarms:

$$POD = \frac{\text{hits}}{\text{hits + misses}} = \text{Probability of Detection}$$

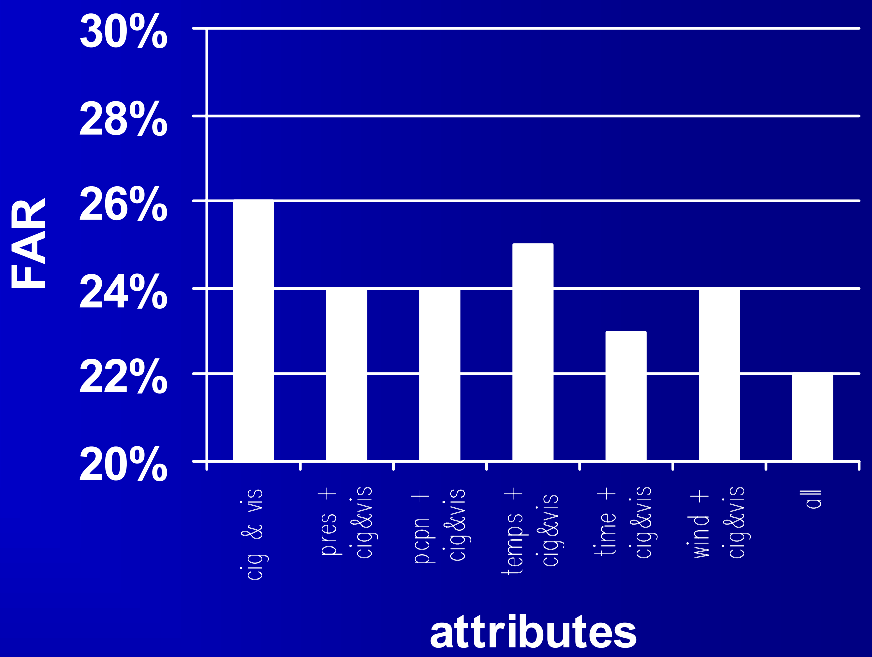
$$FAR = \frac{\text{false alarms}}{\text{hits + false alarms}} = \text{False Alarm Ratio}$$

High POD and low FAR $\Rightarrow$ High accuracy

Experiment 1 - Varying Attribute Set *



POD of "below alternate"

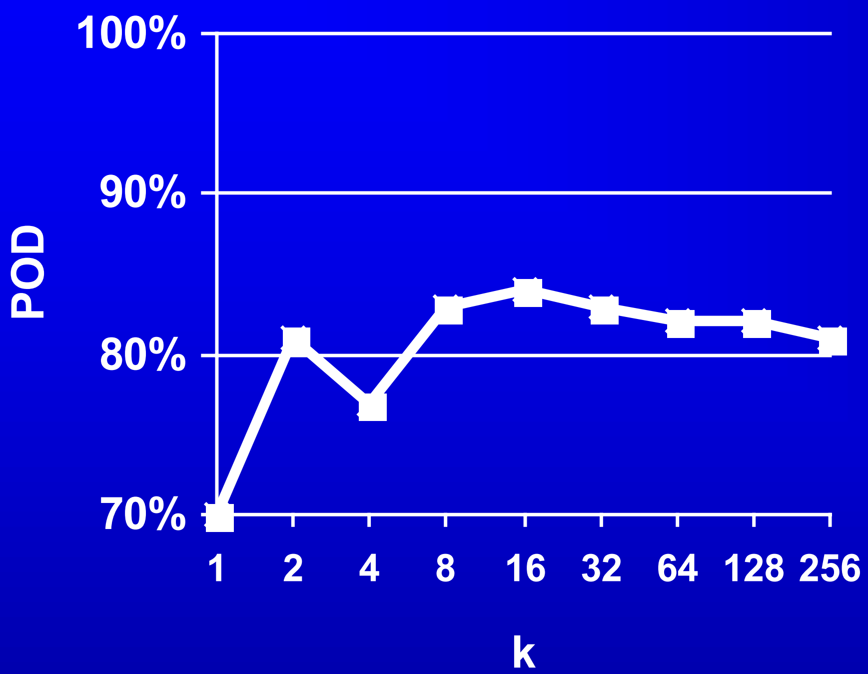


FAR of "below alternate"

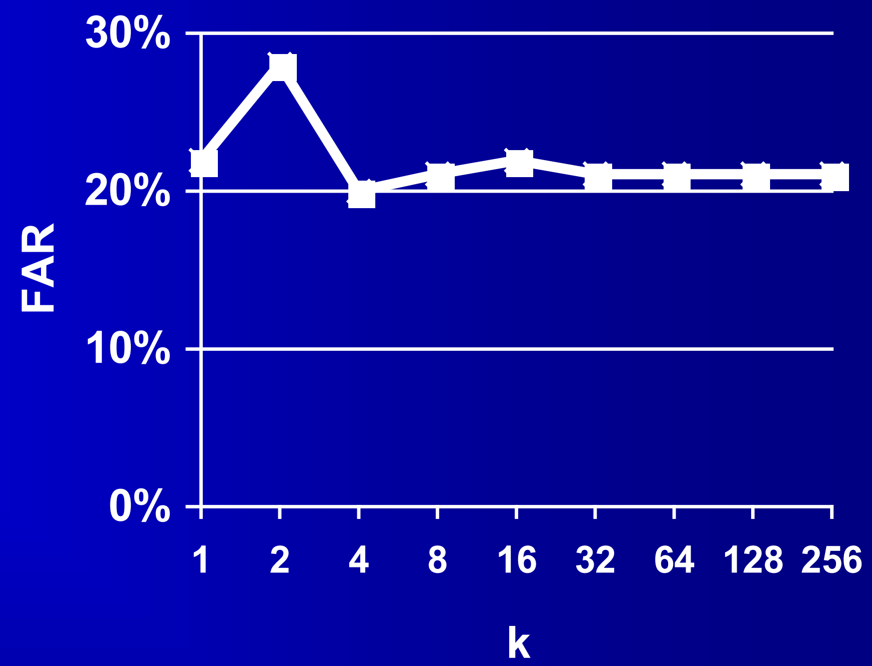
The more attributes used for matching, the more accurate the predictions.

* Seven attribute sets are tested: cloud ceiling and visibility (cig & vis) alone, pressure and cig & vis, precipitation and cig & vis, temperatures (dry bulb and dew point) and cig & vis, temporal attributes and cig & vis, wind and cig & vis, and all of the aforementioned.

Experiment 2 - Varying *k*



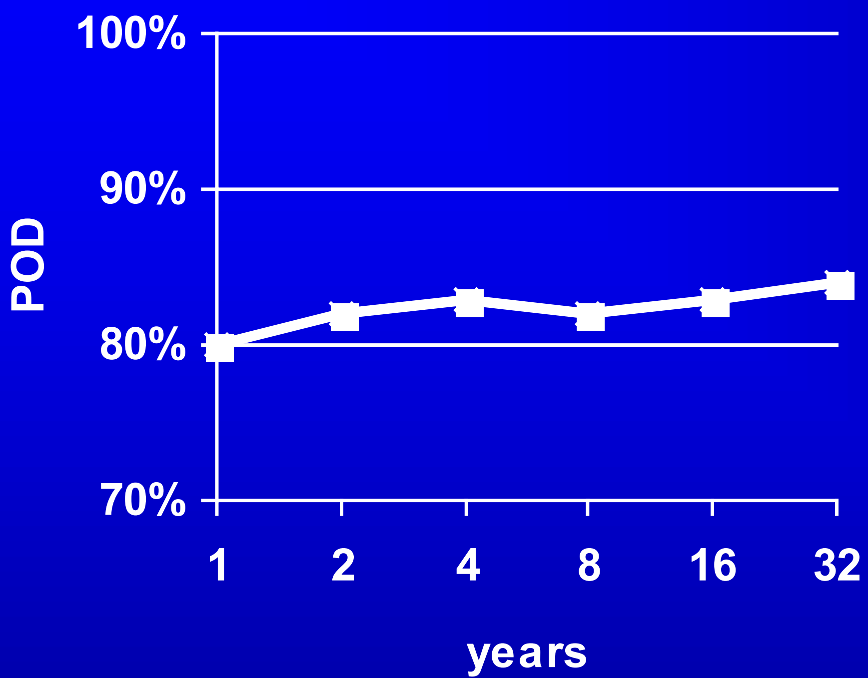
POD of “below alternate”



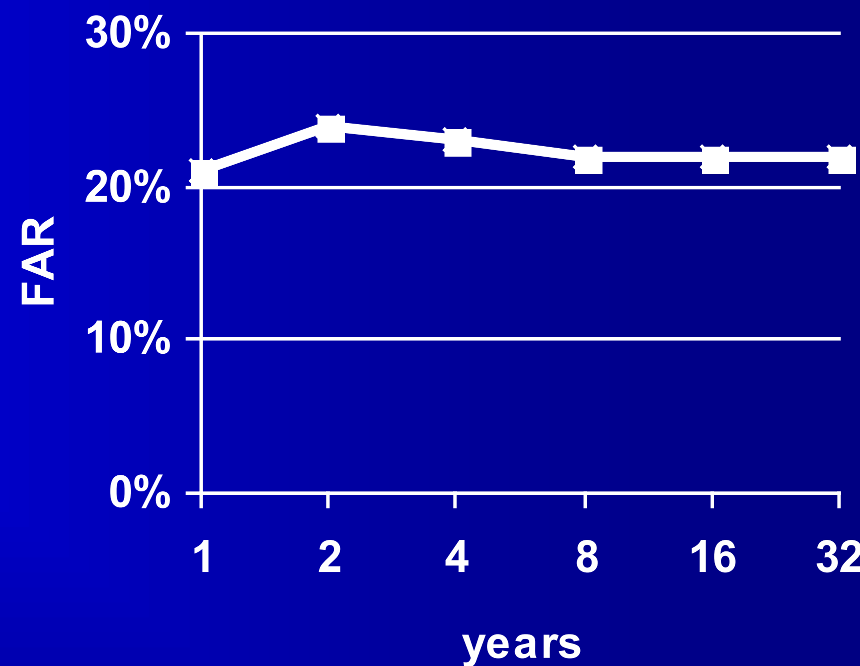
FAR of “below alternate”

Fuzzy k-nn algorithm effective at identifying and ranking *k*-nn. On average, using the 16 nearest neighbors results in more accurate predictions than by using the 256 nearest neighbors.

Experiment 3 - Varying Size of Case Base



POD of “below alternate”

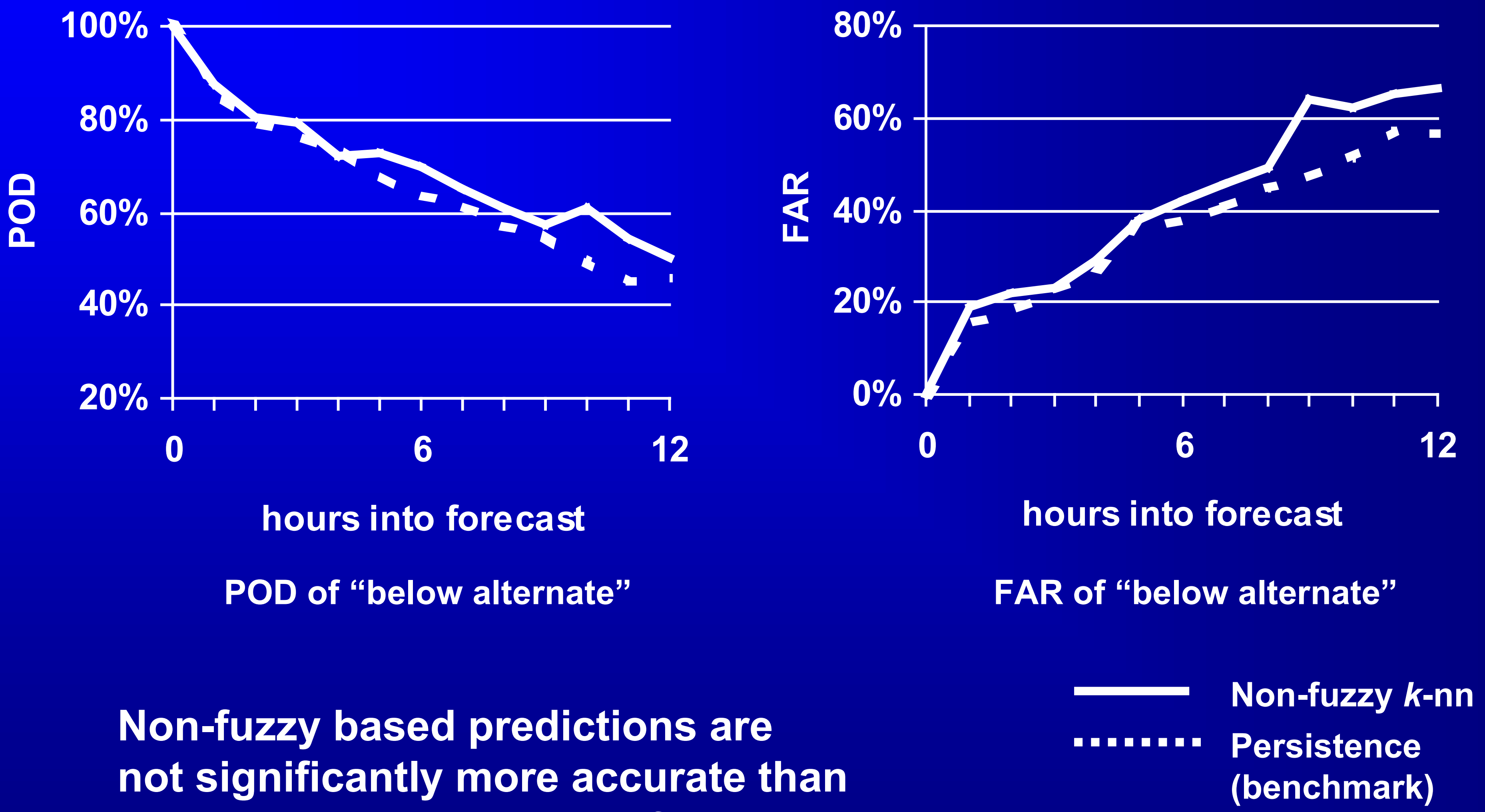


FAR of “below alternate”

Accuracy increases as the size of the case base size increases.

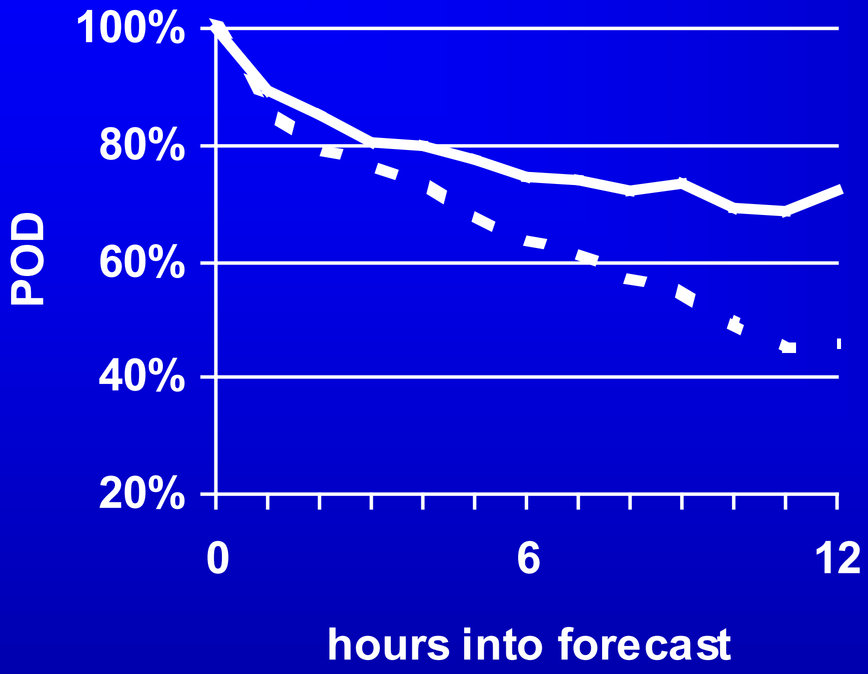
However, relatively high accuracy with a case base size of 4 years means system could be useful for predicting at airports with even relatively small weather archives – most airports.

Experiment 4 - Varying Membership Function

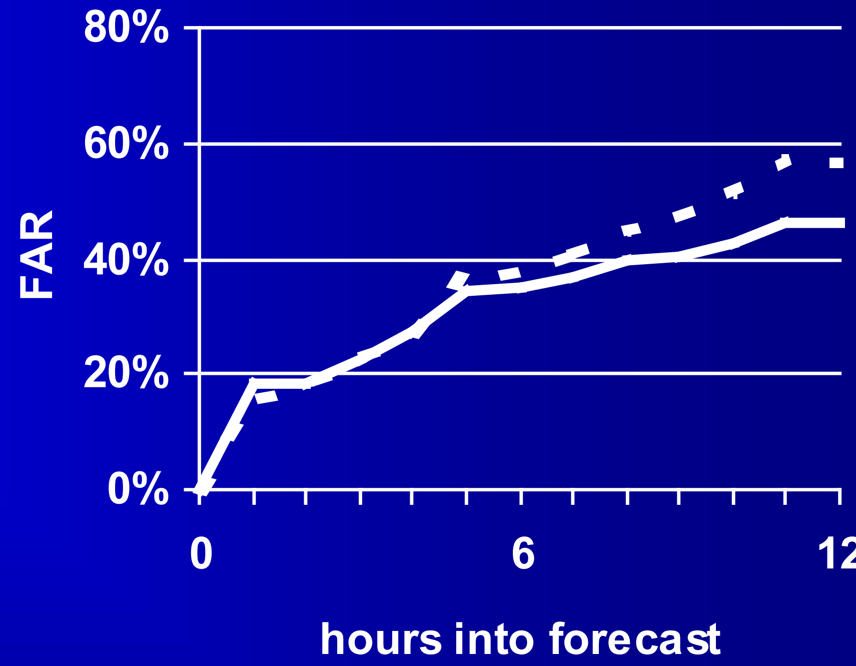


Non-fuzzy based predictions are not significantly more accurate than simple persistence based forecasts.

Experiment 5 - WIND-1 vs. Persistence



POD of “below alternate”



FAR of “below alternate”

Fuzzy *k*-nn

Persistence (benchmark)

Fuzzy *k*-nn based predictions are significantly more accurate than non-fuzzy set based predictions.

Conclusion

Proposed, implemented, and tested a fuzzy *k*-nn based prediction system, called WIND-1.

Its unique component is an expertly-tuned fuzzy *k*-nn algorithm with a temporal dimension.

We tested it with the problem of producing 6-hour predictions of cloud ceiling and visibility at an airport given a database of over 300,000 consecutive hourly airport weather observations.

Prediction accuracy significantly more accurate than benchmark technique and more accurate than non-fuzzy set based technique.

Future Work

Refine WIND system through collaborative R&D with Met Service and academic AI groups.

Develop interface to let users specify special characteristic conditions of present cases to query the data base (e.g., wind shift to west 3 hours hence) and thereby improve the analog selection.

Use computer vision techniques to provide additional predictive information, such projections of weather radar images of precipitation, and projections of satellite images of cloud.

Pursue related research of computer vision and chaotic systems.