
Automatic generation of weather information

Ehud Reiter

(Univ of Aberdeen and Data2Text Ltd)

<http://www.csd.abdn.ac.uk/~ereiter>

Contents

- Background
- SumTime: offshore oil marine forecasts
 - » Issue: word choice, esp. time phrases
- RoadSafe: road maintenance forecasts
 - » Issue: spatial descriptions
- Data2Text: spinout company
 - » Issue: motion
- Conclusion

Background

- Computer programs can automatically produce textual weather forecasts from NWP data
- Many systems deployed for simple forecasts
- Focus here on using AI&NLP technology for forecast generation
 - » Enhance quality
 - » Generate more complex forecasts

FoG System

- Pioneering system in this area built in early 1990s, by Environment Canada and academics from Montreal
 - » Marine forecasts (?)
 - » English and French

Research at Aberdeen

- SumTime: generating simple marine forecasts of high quality
 - » Users prefer to human forecasts in some cases
 - » Word choice better, more consistent
- RoadSafe: generating complex road maintenance forecasts
 - » Again users rate highly
 - » Computer can examine all the data, people cannot
- Data2Text: commercial spinout

Data-to-text

- Our weather forecast generators are a type of *data-to-text* system
 - » Generate textual summary of non-linguistic data
- Draws on artificial intelligence and natural language processing

Other Data-to-Text

- Medical: summarise clinical data for doctors, nurses, patients
- Engineering: summarise operational data for complex machinery
- Assistive technology: summarise statistical data for blind people
- Etc, etc

Language \leftrightarrow World

- Underlying scientific issue: How does language relate to the non-linguistic world?
 - » What do words mean in terms of non-linguistic data?
- Deep question in psychology, linguistics, cognitive science, artificial intelligence, etc

Example questions

- What does *pink* mean in terms of RGB?
- What does *by evening* mean in terms of actual clock time?
- Which shapes are *spikes*?
- Which motions can be called *drifting*?

SumTime

- Generates marine forecasts for offshore oil rigs, from NWP data
- Used for a few years
 - » Forecasters post-edited computer forecasts before releasing them
 - » Worked well, but support a problem
 - Academics poorly suited to software support
 - One reason for spinout (years later)

Example

FORECAST 6 - 24 GMT, Wed 12-Jun 2002

WIND(KTS)

10M: W 8-13 backing SW by mid afternoon and S 10-15 by midnight.

50M: W 10-15 backing SW by mid afternoon and S 13-18 by midnight.

WAVES(M)

...

WEATHER: Mainly cloudy with light rain showers becoming overcast around midnight.

VIS(NM): Greater than 10.

AIR TEMP(C): 8-10 rising 9-11 around midnight.

CLOUD(OKTAS/FT): 4-6 ST/SC 400-600 lifting 6-8 ST/SC 700-900 around midnight

Example (wind speed)

Date	hour	dir	spd
25-10-00	0	SSW	12
25-10-00	3	SSE	11
25-10-00	6	ESE	18
25-10-00	9	ESE	16
25-10-00	12	E	15
25-10-00	15	ENE	15
25-10-00	18	ENE	18
25-10-00	21	NNE	20
26-10-00	0	NNW	26

SSW 12-16 BACKING ESE 16-20 IN THE MORNING,
BACKING NE EARLY AFTERNOON THEN NNW
24-28 LATE EVENING

Steps in summarising data

Standard NLG pipeline

- *Content determination*: deciding what information to include in summary
- *Microplanning*: choosing words, creating sentence structures
- *Realisation*: produce text in weather sublanguage

Content Determination

- Decide on content of forecast
- Use standard data-analysis algorithms
 - » Linear segmentation
 - » Pattern, shape detection
- Adjust for communication
 - » Focus on changes that impact users
 - Eg, affect operational decisions on the rigs
 - » Imitate style of human forecasts
 - Eg, Don't mention flat segments

Choosing words

- Which words used to express the data?
 - » General issue in weather forecasts
- Initially examined what people did
 - » Parallel text-data corpora
 - 1045 (data, human forecast) sets (5 authors)
 - » Post-edit corpora
 - 2728 (data, computer forecast, post-edited forecast) sets (9 authors)
 - » Forecaster suggestions

Time Phrases

Objective: learn numeric time (meaning) of time phrases, from parallel text-data corpus

- 1) Break up sentence into phrases
 - » SSW 12-16
 - » BACKING ESE 16-20 IN THE MORNING,
 - » BACKING NE EARLY AFTERNOON
 - » THEN NNW 24-28 LATE EVENING
- 2) Align phrases with parallel data files
- 3) Statistical analysis of which phrases associated with which times

Alignment

Date	hour	dir	spd	
25-10-00	0	SSW	12	SSW 12-16
25-10-00	3	SSE	11	
25-10-00	6	ESE	18	BACKING ESE 16-20 <u>IN THE MORNING</u>
25-10-00	9	ESE	16	
25-10-00	12	E	15	
25-10-00	15	ENE	15	
25-10-00	18	ENE	18	
25-10-00	21	NNE	20	
26-10-00	0	NNW	26	NNW 24-28 <u>LATE EVENING</u>

Cannot align: BACKING NE EARLY AFTERNOON

Analysis

- Extract (time phrase, time) pairs
 - » (0600, IN THE MORNING)
 - » (0000, LATE EVENING)
- Repeat this exercise for all forecasts in corpus, statistically analyse result

Results: Individual variation

- Some phrases unambiguous
 - » BY MIDDAY
- Some phrases used differently by different forecasters
 - » *BY EVENING*
- Some phrases have same mode, but different distributions around mode
 - » BY LATE EVENING

What does *By evening* mean?

time	F1	F2	F3	F4	F5
1800	30	5	2	27	13
2100	13	6	8	2	11
0000	2	9	80	5	14
other	2	2	1	0	4

Differences significant at $p < .001$

Context-Dependent?

- Does meaning of “by evening” depend on
 - » Season (sunset time)?
 - » Country/culture (typical dinner time)?
- Forecasters said no
- Some users said yes

Verb example

- Forecaster comments
- I prefer *decreasing* to *easing*, as do many others
- I tend to think of *easing* being associated with a slower decrease and/or perhaps with lower wind speeds or heights
- *Easing* is used when trying to indicate a slight decrease when conditions are bad. I.e. getting slightly better (In other words conditions would need to be bad before you would use this).
 - » Contradicts previous (*easing* with low or high wind speeds?)

Terminology Standards

- Standards for usage existed, but not that effective
 - » Sometimes vague
 - » Often ignored
 - » Inconsistent with standards for other types of forecasts read by same users

Our strategy

- Original idea was to try to imitate the word choice of forecasters, gave up on this
 - » Wanted words whose meaning was clear and unambiguous to readers
- Instead looked for words whose meanings were relatively fixed
 - » *by early evening* instead of *by evening* for 1800

Evaluation

- *Question*: how do SumTime texts compare to human-written texts?
- *Method*: ask humans to compare
 - » human-written weather forecast
 - » SumTime-generated weather forecast
 - » ST-Hybrid text: we parse human text, feed output into ST microplanner and realiser
 - Human determines content, but computer chooses words and sentences

Example of variants

Human: S'LY 15-20 BECOMING 22-28 BY THIS EVENING. LATER VEERING S-SW 18-22

SumTime-Hybrid: S 15-20 BACKING 22-28 BY EARLY EVENING, THEN VEERING SSW 18-22 BY MIDNIGHT

SumTime: S 16-21 BACKING SSE 21-26 BY MID AFTERNOON, THEN VEERING S BY EARLY EVENING AND SSW 18-23 BY MIDNIGHT

Subjects

- 72 experienced forecast users
 - » Mostly people who read 100+ forecasts per year, and had been doing so for 10+ years
- Filled out questionnaire when convenient, not at university
 - » Eg, while on boat going out to rigs

Tasks

1. Answer comprehension questions about forecast variant
2. State which variant is better
3. General qualitative comments

Comprehension

Type	Mistakes (lower is better)
Human	4.23
SumTime	3.43
Hybrid	2.11

Pref: SumTime vs Human

	SumTime	Human	Same	P value
More appropriate	43%	27%	20%	0.021
More accurate	51%	33%	15%	0.011
Easier to read	41%	36%	23%	>0.1

Pref: Hybrid vs Human

	Hybrid	Human	Same	P value
More appropriate	38%	28%	34%	>0.1
More accurate	45%	36%	19%	>0.1
Easier to read	51%	17%	33%	<0.0001

Qualitative Comments

- Subjects complained about the usage of ambiguous words in human texts, such as *by evening*
 - » Complained-about words mostly were the words we had identified as meaning different things to different people in our corpus work (my talk last year)

Qualitative comments

- Complained about *later*
 - » Oil-rig forecasters use *later* unambiguously to mean “end of forecast period”
 - » However, shipping forecasters (different people) use *later* to mean after 12 hours
 - In some official terminologies for shipping fcast
 - » Subjects want consistency
 - Terminologies for very specific genre of limited use?

Summary

- Human forecast users think SumTime's computer-generated texts are better than human-written forecast texts
 - » Primarily at word/sentence level
 - » Less difference at content level
- I.e., computers can write better sentences (in forecasts) than people.

RoadSafe

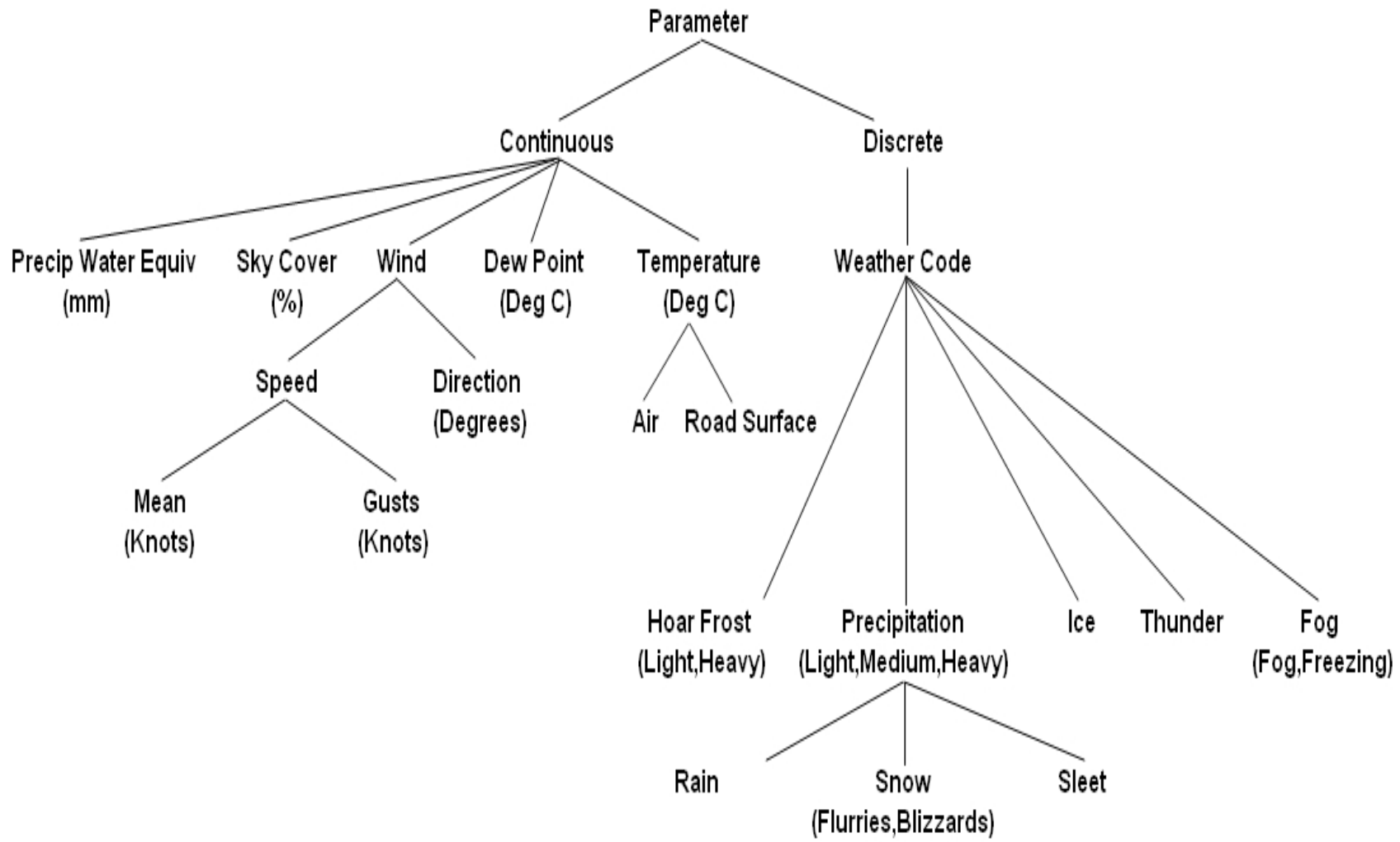


- Generate weather forecasts required by local councils for winter road maintenance:
 - Maximise road safety
 - Minimise environmental impact of excess salting
 - Reduce costs

RoadSafe vs SumTime

- Summary over region, not single point
- Many more time points
 - » Every 20 min vs every 3 hours
- Much more data
 - » 10s of MB vs 1s of KB
- Similar meteorological parameters
- Similar need to focus on what is important to the client

Input: multiple weather parameters



Input: spatio-temporal reference



Human Authored Text

Wind(mph) ...

Weather Cloudy with scattered showers or patchy rain dying out in the evening with some clear spells but some light showers or patchy rain will return late night to early morning. It will become partly cloudy and dry with sunny spells in the morning. The cloud will keep road temperatures above freezing.

RoadSafe Text

Overview Road surface temperatures will reach marginal levels on most routes from this evening until tomorrow morning.

Wind (mph) ...

Weather Light rain will affect all routes this afternoon, clearing by 17:00. Fog will affect some central and southern routes after midnight until early morning and light rain will return to all routes. Road surface temperatures will fall slowly during this afternoon until tonight, reaching marginal levels in some places above 200M by 17:00.

Issue: spatial reference

- How do we describe regions?
 - » *in some places above 200M*
 - » *some central and southern routes*

Frame of reference

- How do we describe a set of points as a region?
 - » Compass: *central and southern*
 - » Altitude: *above 200M*
 - » Coastal proximity, urban/rural, landmarks
- Which specific description is best
 - » False positives vs false negatives?

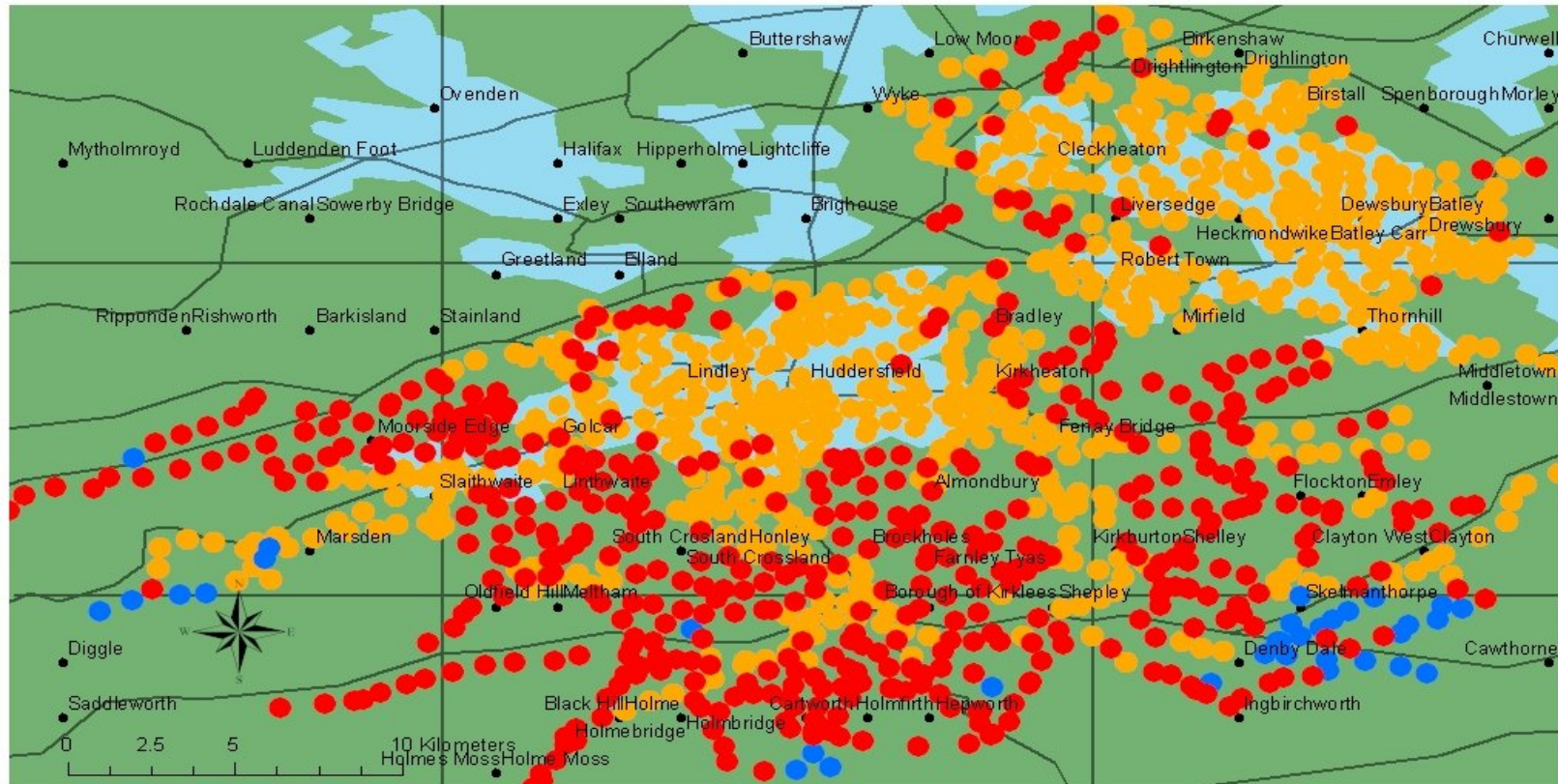
Frame of reference

- Descriptions must make sense causally
 - » Unemployment is high in rural area
 - » Unemployment is high above 200m (??)
 - » Wind speeds are high in rural areas (??)
 - » Wind speeds are high above 200m

First stage: Data Analysis

- **Geocharacterisation**
 - » Using GIS data to add frame of reference information
 - » Altitude, coastal proximity, etc
- **Classification**
 - » Classify non-spatial attributes
- **Clustering**
 - » Group data points based on frame of ref
- **Time series analysis**

Classification



Road Surface Temperatures at 21:00 22/03/08

- marginal
- near critical
- zero

Clustering over time using multiple spatial frames of reference

Parameter: RST	Time	12:00	12:20	12:40	13:00
	Max series (°C)	7.5	7.6	7.7	7.8
	Mean series (°C)	5	5.1	5.2	5.3
	Min series (°C)	2.2	2.1	2.1	2.2
Classification	OK (> 4 °C)				
Density estimate	(proportion)					
Altitude	0M:	1	1	1	1
	100M:	1	1	1	1
	200M:	0.83	0.94	0.96	0.99
	300M:	0.28	0.29	0.34	0.57
	400M:	0	0	0	0
	500M:	0	0	0	0
Direction	CentralNE:	1	1	1	1
	CentralNW:	1	1	1	1
	CentralSE:	1	1	1	1
	CentralSW:	0.95	0.97	0.98	1
	EastNorthEast:	1	1	1	1
	EastSouthEast:	0.98	0.98	1	1
	NorthNorthEast:	1	1	1	1
	SouthSouthEast:	0.7	0.74	0.75	0.87
	SouthSouthWest:	0.46	0.52	0.54	0.64
	TrueNorthEast:	1	1	1	1
	TrueSouthEast:	0.95	1	1	1
	TrueSouthWest:	0	0	0	0
Population	WestSouthWest:	0.56	0.59	0.7	0.72
	Rural:	0.86	0.88	0.89	0.92
	Urban:	1	1	1	1

False positive/negative

- All points above 400M: marginal RST
- All points below 200M: OK RST
- Points 200-400M: some OK, some marginal
- What description to use?
 - » RST marginal above 400M (*false negatives*)
 - » RST marginal above 300M
 - » RST marginal above 200M (*false positives*)
- Relative cost of over/under gritting

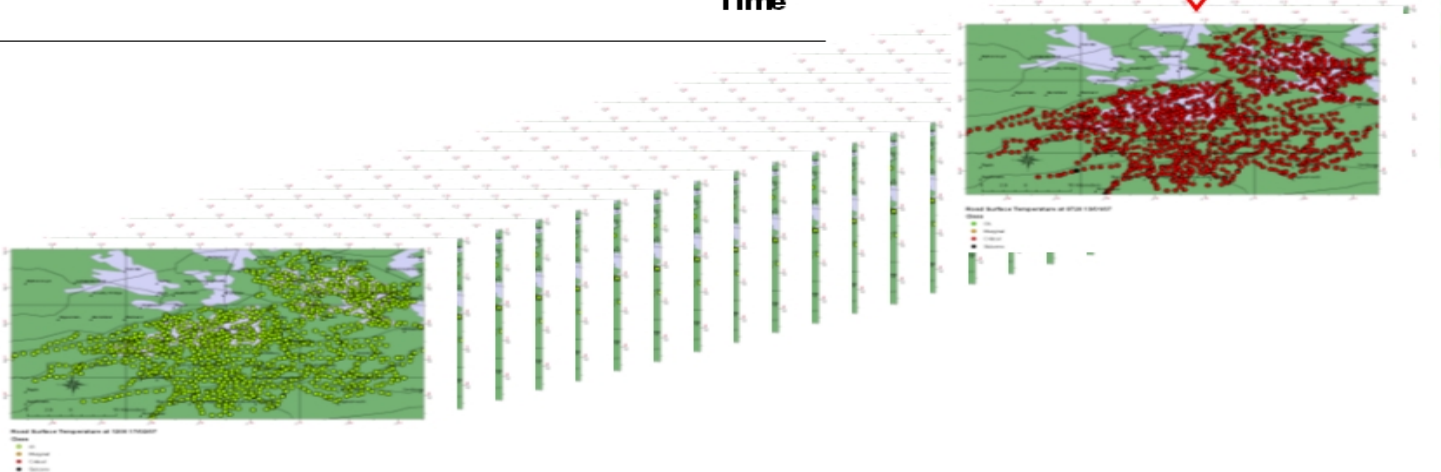
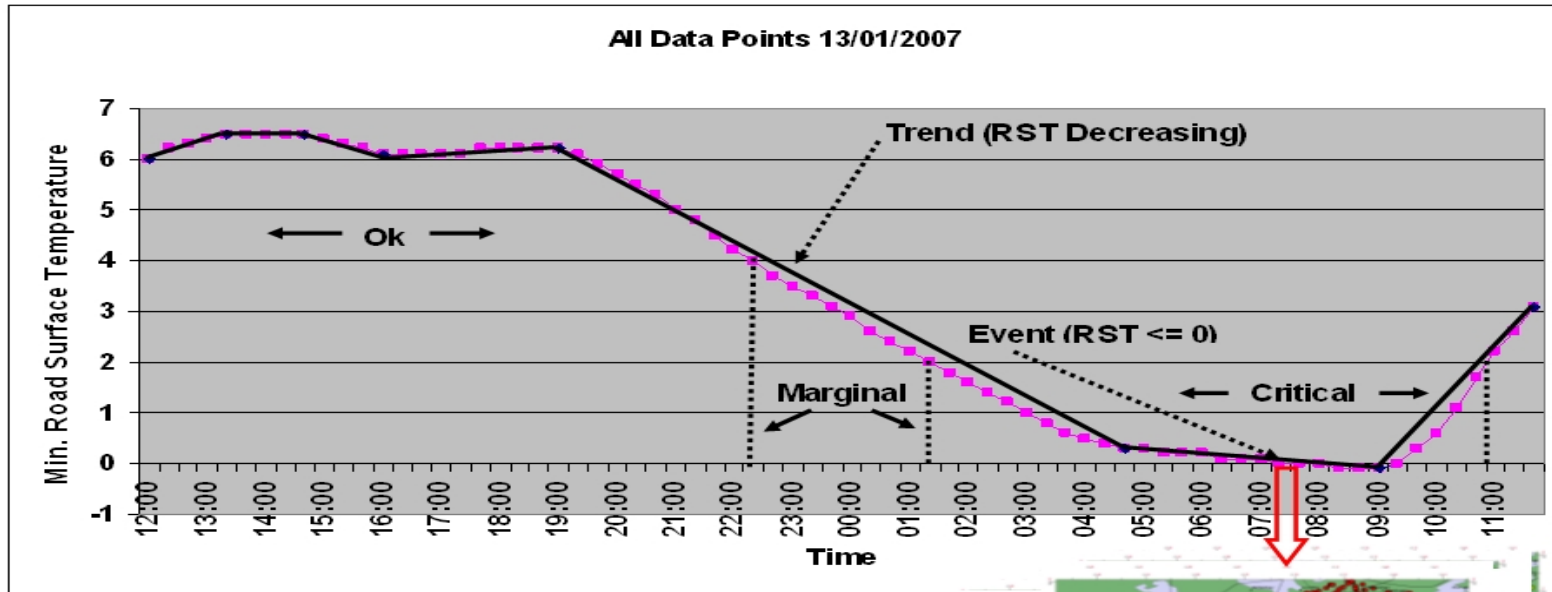
Spatial reference

- How describe sets of points as complex spatial regions in language?
 - » Frame of reference
 - » False positive/negative
 - » Other factors (eg, user expectations)
- Interesting topic, little research

Second stage: Data Interpretation

- Detect important events occurring in space and time
 - » When/where did something important change?
- Based on Keyframe extraction approach (computer vision) and expert knowledge

Data Interpretation



Third Processing Stage: Document Planning

- Select important events to be communicated
- Create structure of the output text
- Rules based upon study of human written examples

Fourth Processing Stage: Microplanning and Realisation

- Takes care of linguistic details to produce a grammatical text
 - » Morphology, Orthographic details etc.
- Similar to SumTime

Evaluation

- 21 road maintenance engineers read human and RoadSafe texts, and responded to a questionnaire
 - » More informative: Roadsafe (85%)
 - » Easier to understand: Roadsafe (60%)
 - » More useful: Roadsafe (70%)

Comments

- Roadsafe preferred because gives more detail (spatial and temporal desc)
- Partially humans are deliberately vague since NWP may be wrong
- But also computer can examine every data point
 - » Humans cannot examine 1M data values, they look at a few and generalise based on knowledge and experience

Comments

- Complaint: no description of motion in Roadsafe texts

Data2Text

- Commercial spinout company
- Doing some commercial confidential work in met (cannot discuss)
- Also working on motion description (govt grant)
- Also non-meteorological projects

Vision

- Generic software modules for forecasts
 - » Input: standard model outputs (eg, GRIB2)
 - » Data analysis: time series, spatial, spatiotemporal
 - » Data interpretation: framework, needs to be instantiated for specific product
 - » Document planning: generic library to support narratives
 - » Microplanning: time phrases, spatial reference, etc

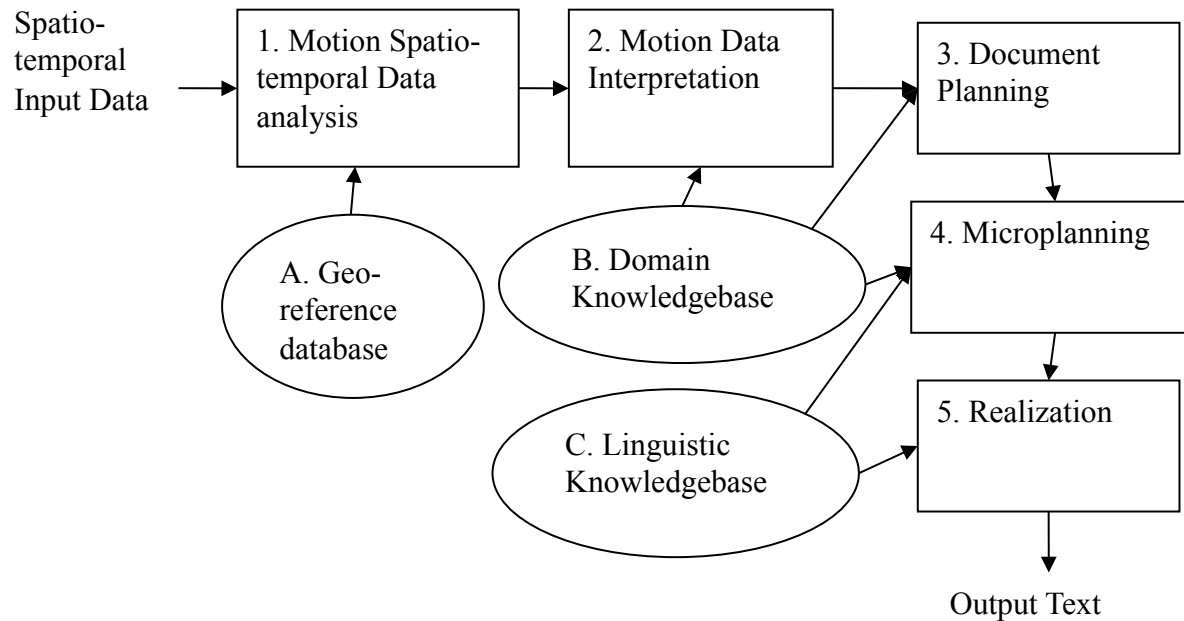
Vision

- System-building methodology
 - » Work with clients to build systems for specific forecast products
 - » Use generic library
 - » Corpus analysis and knowledge acquisition

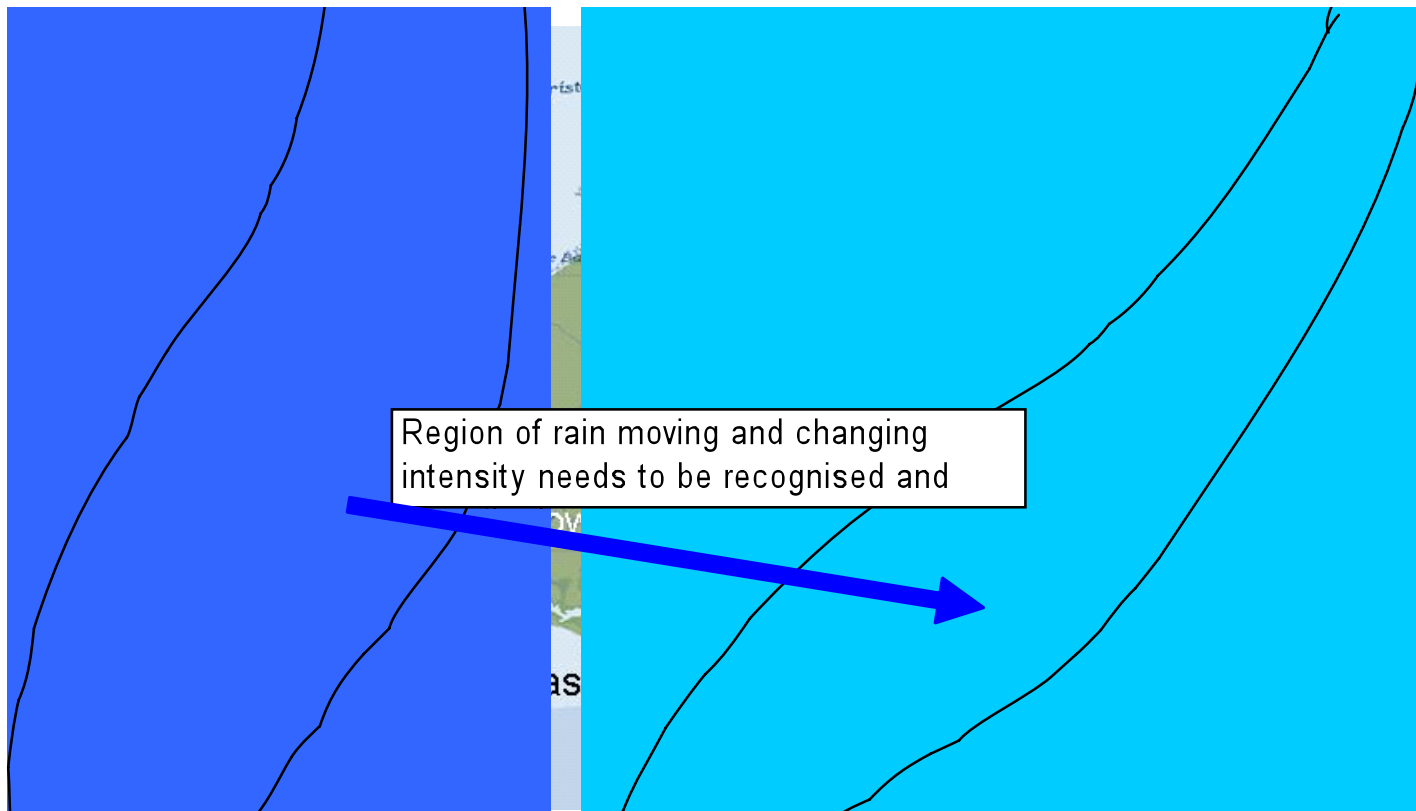
Motion

- Goal: describe moving/changing weather systems
 - » Motion: *reaching the Western Isles by morning and moving quickly east to clear the east coast by late afternoon'*
 - » Motion and change: *A band of rain will approach from the west by 0900, but weaken and die out as it reaches the coast*

SMART motion



Pattern recognition and tracking.



The screenshot shows a web browser window titled "SmartMotion". At the top, there are two dropdown menus: "Select the **area** to run the forecast for:" with "Canada" selected, and "Select the **date** to run the forecast for:" with "Thursday 16 December 2010" selected. Below these is a button labeled "Run the forecast".

24-Hour Forecast for Canada

Valid from 00:00 on Thursday 16 December 2010 to 00:00 on Friday 17 December 2010

Weather Forecast

Isolated outbreaks of light snow probable around 0000 above 500m in the south before patchy snow showers moving in from the south east at 1500, reaching most of the south by the end of the period.

[Data from the NOAA CPC](#) | [Forecast by Data2Text Ltd](#)

Data analysis

- Start with computer vision techniques
 - » Track moving objects in video
- Problem: weather systems change shape (unlike objects in videos)
 - » So need to adapt techniques, cannot use “off the shelf”

Linguistics

- Meaning/usage of words describing motion and change
 - » When should “*weaken*” be used to describe a changing weather system?
- Empirical corpus analysis

Status

- Ongoing project, still in early stages
 - » Initial focus: *moving, spreading* and *clearing*
- Demo

Summary

- **Computer AI/NLG systems can generate high-quality textual forecasts**
 - » Better than human in some cases
 - Better consistency, word choice
 - Computer can examine all the data
- **Many scientific challenges**
 - » Variation in human language, description of spatial regions, motion
- **Also engineering challenges**