Meteorological Bulletin Automatic Generation based on Spatio-Temporal Reasoning

Hua-Ping Zhang	Huan-Ping Wu	Jian Gao	Yan-Ping Zhao	Zhong-Liang Lv
Beijing Institute of Technology	China Meterological Administration	Beijing Institute of Technology	Beijing Institute of Technology	China Meterological Administration

Email: Kevinzhang@bit.edu.cn

Abstract

Meteorological bulletin has more and more diversified, large scale, highly integrated requirements and potential demands from whole society. The strong professional efforts involved in transforming the variety of special meteorological data to natural language text are becoming more challenging in providing sophisticated and easily understood weather features. This paper presents a new Meteorological bulletin automatic generation method based on spatio-temporal reasoning. To enhance an exact and non-redundant description for complex meteorological data, and for special future tendency dynamics in emerged interesting areas. We also evaluate this method with real data from National Meteorological Center and prove that it's feasible and effective after implementing.

1 Introduction

Meteorological bulletin has been, and is more and more exploited in many fields such as for commercial, military, agriculture, particular purposes or potential environment influences. With the increasing demand, variety of meteorological bulletins written with natural language needs huge amount of professional effort to analysis meteorological data and complicated rearrangement for the ordinary users to understand and with special features.

Writing meteorological bulletin to report daily weather is time-consuming and tedious. And there has been some inherit rules behind so many reports. Therefore, it is necessary to improve the ability for automatically generate meteorological bulletin (AGMB).

The development of visualization technology in the field of meteorology has been gradually mature, but automatic or semi-automatic generating technology for text products haven't got fully development due to the difficulties in effective express on the meteorological features in varieties of irregular areas (see the different colored rain circles occupy ambiguous multiple regions in Figure 1).

Figure 1: The three rain circles in Northeast for example

Therefore, we proposed a new method based on spatio-temporal reasoning of AGMB to enhance service variety in text generation and highly integration. And it also raises forecast accuracy and service quality levels.

2 Previous Work

There are few researches since the early 1970s, and most are based on template-based Computer Worded Forecasting (CWF). It establishes a template for meteorological information to be generated, and transform into appropriate words of bulletin. The representatives are IFPS in 1970s, ICWF in 1993 and Siren System in 1999 [1].

Natural Language Generation (NLG) starts in 1960s and has developed fast in recent years. FoG is a representative system in Canada [2-3]. This system generates marine weather forecast text (English or French) after simple data analysis and content structuring. SumTime_Meteo, developed by University of Aberdeen, UK[4]. It collected forecast corpus, and realizes automatically generating forecast text through data analysis, document planning and macro planning.

MLWFA [5] was developed by Shanghai Jiao Tong University, China, since 1999. It took NLG, especially on framework generation, text planning, sentence optimization and text consistency. But there was no further progress in last decade.

We summarize differences between our new research and previous work as following 4 aspects:

(1)**Description of scope**: China is vast geographic area with complex and protean weather phenomenon, other countries have relative small ones.

(2)**Spatio and Time scale**: CWF or NLG, most previous researches focused on phrase organization, but ignored the time and space characters of meteorological data. This paper takes more attention on spatio-temporal reasoning to describe meteorological phenomenon clearly and exactly with details.

(3)Language: Previous systems are mostly designed in English or French, and this paper discusses how to express in Chinese with segment.

(4)**User**: The systems mentioned are mostly used for oil mining to forecast sea weather. But we serve the public, special purpose forecast features.

3 Our New Method

We propose a new method based on spatio-temporal reasoning to establish text framework used to organize and plan the forecast language, and leave some slots to fill into geographic or meteorological information with consistency rules we establish from complex meteorological data.

3.1 Test generation framework

The framework we proposed is made up of 3 parts: phrase planning, sentence planning and chapter structure planning (see Fig 2).

In **phrase planning**, we analyze weather forecast bulletins published by National Meteorological Center in the last two years, a total of 1963 files. After data analysis, we get some meteorological, geographical hierarchy **variables** and a lot of **sentences** with slots.

In **Sentence Planning**, we establish Sentence Template Database according to the machine learning weather **features** from the original GIS data files. They are all in four kinds: snow&rain, wind, temperature and visibility. Then we choose appropriate

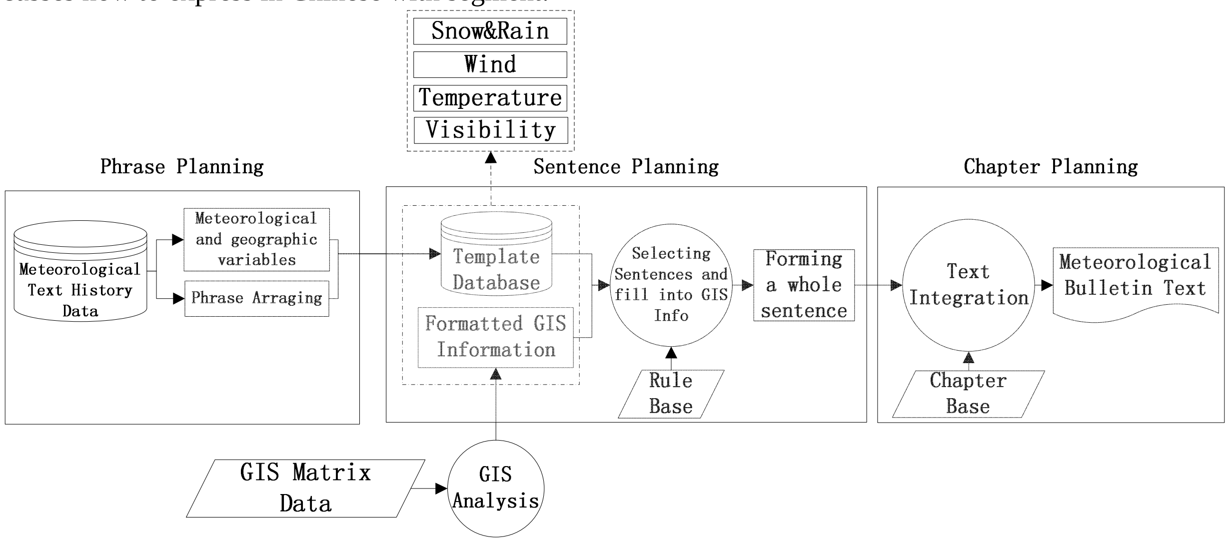

Figure 2: The framework of forecast text generation.

sentence template according to Rule Library, and put the GIS information into the template after spatio-temporal reasoning, and get whole sentence describing weather exactly and simply.

In **chapter structure planning**, we take adjacent words statistics combined with word frequency statistics to extract context structure used by weatherman, and make text integration according to the Chapter Structure Library. Then we gain a whole meteorological bulletin text.

3.2 Variable

In **phrase planning**, we extract some describing weather **variables** as **"小雨"** (light rain), **"中雨"** (middle rain), **"大雨"** (heavy rain), **"暴雨"** (torrential rain) and so on. The **positional variables** describe the position of some weather. In order for clarity, we divide it into 4 hierarchies in Table 1.

Table 1: Geographical variable list

1st	" North China ", "Inner Mongolia", "Yellow River", "South China", "Tibet"
2nd	based on 1st level the western of North China, central part of Inner Mongolia, "江南东部" (the eastern of Yangtsi River)and so on
3rd	Every province: "山东省"(Shandong Province) , "新疆维吾尔自治区" (XinJiang Weiwuer province), 黑龙江 (Heilongjiang province)
4th	Sub area of 2nd or 3rd levels，for example, "山东东部" (the eastern of Shandong Province), "黑龙江北部" (the east to Heilongjiang Province)

There are 3 kinds of **inclusion** relationships in the 4 levels of geographical variables. The 1^{st} level is the biggest area, which includes the 2^{nd} or the 3^{rd} , as North China includes Heilongjiang province; The 2^{nd} or the 3^{rd} includes the 4^{th}, as "Heilongjiang province" includes "the east to the province". But the 2^{nd} and 3^{rd} are mutually exclusive.

3.3 Spatial reasoning for geography

China is so vast and the weather is too complex and protean to describe the geographical phenomenon. We design a range of rules to solve the problem and construct a Geographic Information system (GIS) interface, to gain the **features** and **variables** as semantic XML output shown as Fig 3.

```
<item>
   <id>7</id>
   <geo_name>西北地区</geo_name>
   <proportion>0.010</proportion>
   <area>29176</area>
   <type>雨雪</type>
   <weather_name>暴雨</weather_name>
   <value>50</value>
</item>
```

Figure 3: The rain circle information in XML

We know the weather, the area and calculate the proportion of each feature and each variable by the output, in every 1st level area and 2nd and so on in one day. But, we couldn't make it non-redundant. The middle rain appears in multiple areas repeatedly and separately as follows.

"东北有中雨，东北东部有中雨，东北西部有中雨，黑龙江有中雨"

So we set a range of rules to treat this. We define the rain proportion of an area as ***F***. Then we establish a interacting algorithm: if the area is in 1^{st} level, and ***F*** value is above some limit1, then we output the geo-name and the feature for the position of the area, if it is below some limit, we skip 1^{st} and goto the 2^{nd} or 3^{rd}, and see if the ***F*** in the 2^{nd} or 3^{rd} is above some limit2, if so, then output the position in the level one by one, if it is below the limit2, we also skip the level and goto level 4^{th}, and do the same like for some limit3. The details are showed in Figure 4.

```
if(F<=0.05)                  Ignoring output
else if(F>=0.9)              Output : 西南
else if(0.7<=F<0.9)          Output : 西南大部
else if(F<0.7)//Go into the 2nd and 3rd level
{
   //Judging 2nd or 3rd
   if( IS_Second() )//if go into 2nd
      Position Analysis();    Output : 西南东部和南部（例）
   else If( IS_Third())//if go into 3rd
   {
      //Analyze every privince one by one
      if(f<=0.05)             Ignoring output
      else if(f>=0.9)         Output : 云南（例）
      else if(0.7<=f<0.9)     Output :云南大部（例）
      else if(0.3<f<0.7)// Go into 4th
         Position Analysis();       Output :云南西北部|云
                                            南局部（例）
   }
}
```

Figure 4: The process of geographical analysis.

We also design some rules for some group of 2^{nd} or 3^{rd} areas to describe the geographical position. For example in "Northeast", it could be two possible divided groups: Eastern and Western, or Southern, central part and Northern. If the light rain is shown as Fig1(the light green part), we can describe as either **"东北东部和西部"**(the eastern and western of Northeast) or **"东北南部和北部"** (the northen and southern part of Northeast). Obviously, human can judge that and the latter is preferred but computer not that easily, then we carefully setup some rules for computer to do spatio-position more accurately.

3.4 Tendency analysis

Sometimes, the rain circles in the next 3 days have their trends, so we need to analyze the changing positions of the circles and to determine which direction all circles to move. But the circle today may be divided into 2 parts tomorrow, or several circles today may merge into a whole tomorrow. And the number of circles everyday is often large. So we consider a spatio-temporal reasoning algorithm for all the circles.

In this module, we mark three IDs as the weather, the area and the day of every circle in the GIS interface. We establish a tendency model to simulate how the circles move. In this model, every circle of today has a mapping from every one of yesterday and to every one tomorrow, and the mapping has its property shows the direction from the previous circle to the next ones. The direction has 8 kinds of value: East, West, South, North, Northeast, Northwest, Southeast and Southwest. We can get all the mappings with directions, and then we only leave the mappings whose previous directions are equal to the next one. The details are shown in Figure 5.

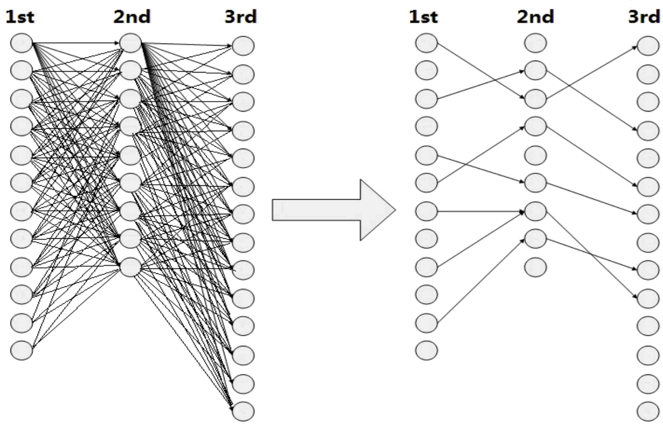

Figure5: The model of mapping circles.

Through dropping the different mapping, we can get limited directions. And at last we will get only one direction after calculating frequency with the weights of every circle's area and weather features. Now, we can fill variable information into the sentence template after tendency analysis to generate the whole sentence describing tendency as following:

未来三天，我国自西向东将有一次小到中雨（雪）天气过程。

4 Experiment

With the help of National Meteorological Center, we develop a system used for weather forecast automatically generation. Then we randomly select manual weather forecast texts and corresponding original GIS data of 180 days from the past meteorological history data, and use the system to automatically generate forecast text according to the GIS data so as to compare two kinds of forecast.

The AGMB forecast texts are manually checked by 3 experts from National Meteorological Center, and experts give 5 levels of score for the strengths of the similarity between two kinds. And at last it's proved that the forecast automatically generated by our system is easy-understood, highly similar and feasible. The detail is shown in Table 2.

Table 2: Evaluation results

Forecaster	Best	Good	Medium	Bad	Worst	Eligibility Rate
Qin Huafeng	116	25	17	19	3	0.87
Rao Xueqin	123	30	10	11	6	0.91
Zhang Wendong	121	18	20	13	8	0.88

5 Conclusion

In this paper, we have established a framework of weather bulletin automatically generation and used spatio-temporal reasoning method to analyze GIS information to generate exact, non-redundant bulletin text with detailed weather features and variables. We also implement and evaluate the system based on the methods for National Meteorological Center, China. There will be a lot of services provided by the departments of the Center. It also means that the potential market of automatically or semi-automatically generation meteorological bulletins is very large. There are still many new fields need to be exploited on bulletin text generation.

Acknowledgments

We thank National Meteorological Center for its GIS technical support and help in project testing.

References

[1]NLG Systems Wiki, Available at: LISThttp://www.nlg-wiki.org/systems/.

[2] E. Goldberg, N. Driedger, and R. Kittredge. Using Natural-Language Processing to Produce Weather Reports[J]. IEEE Expert: Intelligent Systems and Their Applications, Volume 9, Issue 2 (April 1994) pp:45-53.

[3] Eli Goldberg. FOG: Synthesizing Forecast Text Directly from Weather Maps[R]. Artificial Intelligence for Applications, 1993. Proceedings of Ninth Conference, Orlando, FL, USA. Available at: www.aclweb.org/anthology/C/C86/C86-1132.pdf

[4] Somayajulu G. Sripada, Ehud Reiter, Jim Hunter and Jin Yu. SUMTIME-METEO: Parallel Corpus of Naturally Occurring Forecast Texts and Weather Data - Technical Report AUCS/TR0201[R]. Computing Science Dept., University of Aberdeen. Available at: http://www.csd.abdn.ac.uk/~ssripada/sumtime/SumTime_Meteo.rtf

[5] Tianfang Yao, Dongmo Zhang, Qian Wang. MLWFA: A Multilingual Weather Forecast Text Generation System. In Proc. of 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000): Tutorial Abstracts and Demonstration Notes. Hong Kong, China, Oct., 2000, 13~14.

[6] Alfred. V. Aho and Jeffrey D. Ullman. 1972. The Theory of Parsing, Translation and Compiling, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

[7] Ehud Reiter, Somayajulu Sripada, Jim Hunter, Jin Yu, Ian Davy. Choosing words in computer-generated weather forecasts. Artificial Intelligence, 2005, 167(1-2): 137·169.

[8] Anja Belz. Probabilistic Generation ofWeather Forecast Texts. Proceedings of NAACL HLT 2007. pages 164-171.

[9] San-min Wang, Zhao Bin, A logical framework for fuzzy quantifiers Part I: Basic properties，Advances in Soft Computing. 42, 602-611, 2007.

[10] Pani AK, Bhattacharjee GP. Temporal representation and reasoning in artificial intelligence: A review. Mathematical and Computer Modelling, 2001, 34(1-2):55~80.

[11] Cohn AG, Hazarika SM. Qualitative spatial representation and reasoning: An overview. Fundamental Informatics, 2001, 46(1-2): 1~29.

[12] Zhang XF, Cui WH. Spatio-Temporal analysis and modeling based on the integration of GIS and CA model. Journal of Image and Graphics, 2000, 5A(12):1012~1018 (in Chinese).

[13] Zheng KG, Tan SY, Pan YH. A unified spatio-temporal data model based on state and change. Journal of Software, 2001, 12(9): 1360~1365 (in Chinese).