

Analys och prediktering med osäkra data ur det geografiska rummet¹

Johan Schubert
Institutionen för Informationssystemteknik,
Försvarets forskningsanstalt, 172 90 Stockholm
schubert@sto.foa.se
<http://www.foa.se>

Abstract

We give a survey of techniques which were proposed to improve the methodology for interpreting uncertain submarine intelligence reports and making predictions about future positions based on historical and geographical information.

1. Introduction

In 1980, an event occurred which caused considerable military and political commotion in Sweden. For a period of several weeks, the Swedish navy hunted what it later judged to be two foreign submarines operating in the country's inner territorial waters, near Sweden's largest naval base.

This event commenced a more than decade-long period of political uneasiness and increasing military, as well as public, vigilance. The period was characterized by an inflow of final event intelligence reports to the Swedish military headquarters, which during the years of 1986-88 reached a peak of about 1 000 per year. In total, more than 6 000 reports were collected which may refer to independent sightings or detections of foreign underwater activity, [1].

The purpose of this paper is to describe and discuss some aspects of a peculiar class of data analysis applications which arise from bursts of spatio-temporal phenomena, in this case represented by sparse military intelligence data.

In target tracking research, the usual assumption is that the visibility of the target(s) is fairly stable, so that targets can be tracked in real time, perhaps with some ambiguity arising from target multiplicity and/or noisy and fluctuating signals. The situation we describe here is very different: the targets are usually seen only in short glimpses, precise classification or identification is almost never possible, target location is often poorly defined, false alarms occur frequently (some say always), and the "sensors", including human observers, are of many different kinds, each with its own special characteristics.

In Sec. 2 we survey a proposed method for making short-term predictions about future positions based on a combination of historical and geographical information, [2].

2. Making Tactical Predictions from Learned Patterns

The idea is to develop methods which recognize an incoming sequence of intelligence reports as belonging to a certain category of sequences. Having found such a category of sequences we obtain probabilities for different future developments given the current situation.

¹Based on a paper at the *Third Int. Conf. on Knowledge Discovery and Data Mining (KDD-97)*, [3].

Our intelligent system learns the categories of sequences of (simulated) intelligence reports, Fig. 1. This learning is done by genetic algorithms. When we receive a new scenario it is analyzed using the learned categories. If the system finds a category of sequences with a beginning similar to the current sequence, the remainder of the historical sequences are used to give a prediction about the future, Fig. 2. The result of the prediction is compared with the future of the scenario. During the learning of prediction rules, statistics are gathered about how often every prediction rule makes a correct prediction and also about the accuracy of the prediction.

To be able to learn from data we must first find a suitable way to represent the problem. The region under study is a 50 x 50 square kilometers area in the southern Stockholm archipelago. We have chosen to divide this area into elements of 1 x 1 square kilometers. Since we want to make predictions from the time of a particular report we use a relative time scale starting from the current report's indication time. In learning and prediction we will only consider reports which occurred during the last 24 hours. These 24 hours are represented by 120 time intervals of twelve minutes each. Such a representation is shown in Fig. 3.

We have here a sequence of three intelligence reports. The latest report from area E5 is placed to the right on the time scale at T0, and the two earlier reports from area I4 and F4 are placed in their respective time intervals, T5 and T3.

In learning we will now try to predict the next event. Suppose it will happen in area E7 two time intervals into the future, Fig. 4. Our aim is to find a rule that predicts this event based on the earlier information. Such a rule can be very specific in both precedent and prediction:

If [I4 & T5] [F4 & T3] [E5 & T0] then [E7 & T-2],

or it may throw a much wider net, e.g.:

If [HIJ345 & T456] [EF34 & T1234] [DEF456 & T0] then [CDEF67 & T-1-2], Fig. 5.

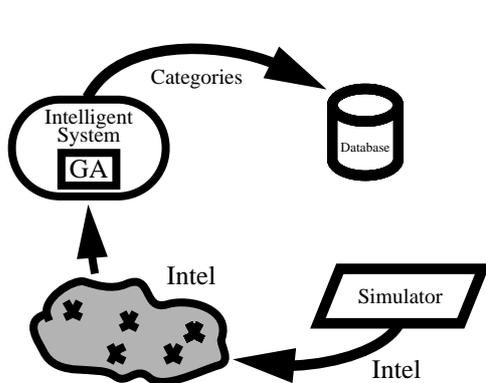


Fig. 1. The system learns categories of sequences of intelligence reports.

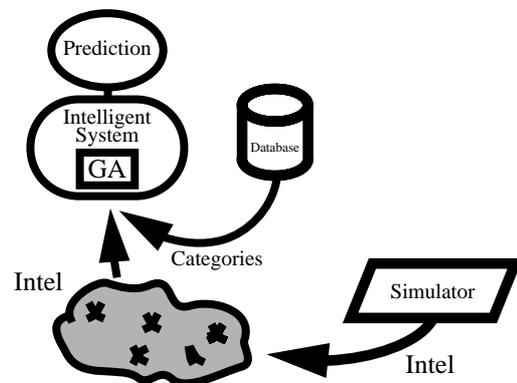


Fig. 2. The system makes a prediction based on a new sequences of intelligence reports.

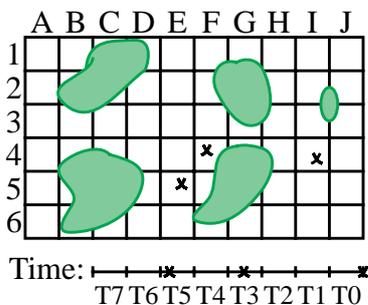


Fig. 3. Representation.

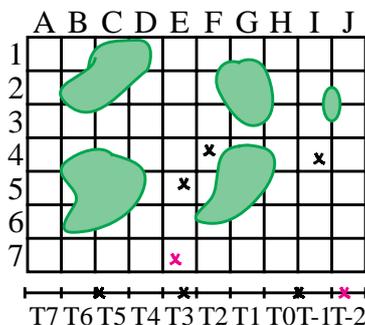


Fig. 4. Find a rule to predict the next report.

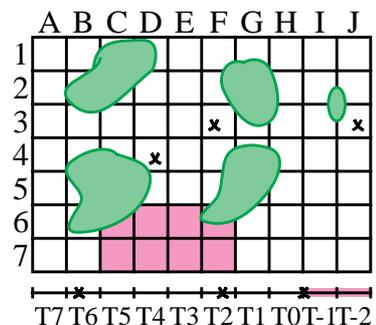


Fig. 5. A prediction rule makes a prediction about an area and a time interval.

Both alternatives and all other combinations with a certain specificity in the precedent, or parts of it, and another specificity in the prediction are possible and automatically tested during the learning phase. The advantage of a specific prediction is obvious, but a disadvantage may be that the prediction rule tends to become a special case and may also get a low probability. A less specific rule has a higher probability but is not as useful in the individual case. The learning mechanism uses a scoring method that takes this into account and finds a suitable balance. During the learning phase, prediction rules are automatically generated from earlier successful rules. In this way ever more successful rules may be generated until a final optimum is reached. The remaining rules are evaluated statistically and will henceforth provide decision support.

A statistical analysis based on a simulation of the method showed that the probability of a correct prediction was at best 54%, with an accuracy in predicted position of 5 kilometers and in predicted time of 48 minutes. The southern Stockholm archipelago, the area used in these trials, covers about 50 x 50 square kilometers. A submarine hunting operation usually lasts several days. In this situation, prediction rules with a probability and an accuracy such as these should be very useful.

References

1. *Ubåtsfrågan 1981-1994. Rapport från Ubåtskommissionen.* Statens Offentliga Utredningar 1995:135, Stockholm 1995.
2. Schubert, J., Prediktion av främmande ubåtars framtida position, FOA-R--96-00296-3.4--SE, Avdelningen för Ledningssystemteknik, Försvarets forskningsanstalt, Linköping, 1996.
3. Bergsten, U., Schubert, J., and Svensson, P., Applying Data Mining and Machine Learning Techniques to Submarine Intelligence Analysis, *Proc. Third Int. Conf. on Knowledge Discovery and Data Mining*, Newport Beach, August 14-17, 1997.