Prognostic Model for Early Warning of Threatening Influenza Waves

Rainer Schmidt, Lothar Gierl

Institut für Medizinische Informatik und Biometrie,
Universität Rostock
Rembrandtstr. 16 / 17
D-18055 Rostock
[rainer.schmidt / lothar.gierl] @medizin.uni-rostock.de

Abstract: The goal of the TeCoMed project is to send early warnings against forthcoming waves or even epidemics of infectious diseases, especially of influenza, to interested practitioners, pharmacists etc. in the German federal state of Mecklenburg-Western Pomerania. The forecast of these waves is based on written confirmations of unfitness for work of the main German health insurance company. Since influenza waves are difficult to predict because of their cyclic but not regular behaviour, statistical methods based on the computation of mean values are not helpful. Instead, we have developed a prognostic model that makes use of similar former courses. Our method combines Case-based Reasoning with Temporal Abstraction to decide whether early warning is appropriate. In this paper, we present this method.

1. Introduction

Many people believe influenza to be rather harmless. However, every year influenza virus attacks worldwide over 100 million people [Ni95] and kills alone in the United States between 20.000 and 40.000 people [Hw99]. The most lethal outbreak ever, the Spanish Flu in 1918, claimed 20-40 million lives worldwide, which is more than the second world war on both sides together [Do83]. So, influenza is the last of the classic plagues of the past which has yet to be brought under control and consequently, in the recent years many of the developed have started to generate influenza surveillance systems countries (e.g. US: www.flustar.com, Japan [Sh01] and France [Pr01]).

The goal of our TeCoMed project is much more humble, namely to send early warnings against forthcoming waves or even epidemics of infectious diseases, especially of influenza, to interested practitioners, pharmacists etc. in the German federal state Mecklenburg-Western Pomerania. Available data are written confirmations of unfitness for work, which have to be sent by affected employees to their employers and to their health insurance companies. These confirmations contain the diagnoses made by their doctors. Since 1997 we receive data from the main German health insurance company. Figure 1 shows the courses of the last four influenza periods (from October to March) for Mecklenburg-Western Pomerania.

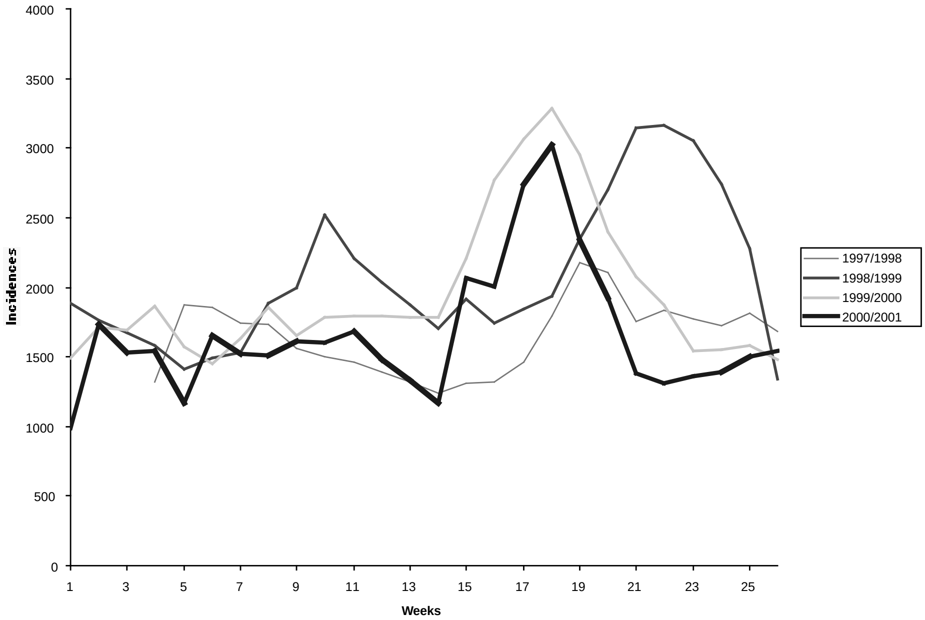

Figure 1. Influenza courses for Mecklenburg-Western Pomerania from October till March. The 1^{st} week corresponds to the 40^{th} week of the calendar and 14^{th} week to the 1^{st} week of the next year.

Usually, each winter one influenza wave can be observed in Germany. However, the intensities of these waves vary very much. In some years they are nearly unnoticeable (e.g. in the winter of 1997/98), while in other years doctors and pharmacists even run out of vaccine, which in Germany last occurred in December 1995. Furthermore, figure 1 shows that the influenza waves occurred in February and March. However, we know that this is probably accidental and in Germany a wave may already start in December.

Influenza waves are difficult to predict, because they are cyclic, but not regular [FB97]. Because of the irregular cyclic behaviour it is insufficient to determine average values based on former years and to give warnings as soon as such values are noticeably overstepped. So, we have developed a method that combines Temporal Abstraction [Sh97] with Case-based Reasoning [Le98]. The idea is to search for former, similar cases and to make use of them for the decision whether early warning is appropriate.
Viboud et al. [Vi01] apply the method of analogues [Lo69], which originally was developed for weather forecasting. It also takes former, similar courses into account. However, the continuations of the most similar former courses are used to predict

future values, e.g. the influenza incidences of next week. Instead, we intend to discover threatening influenza waves in advance and to provide early warnings against them.

2. Methods

Within the ICONS project we have already developed an early warning system, namely for the kidney function [SPG99], which we have presented in the CBR community [Sc96]. Our method concerning the kidney function combines Case-based Reasoning with Temporal Abstraction. For predicting influenza waves, we apply the same ideas and methods again. Since CBR is well-known to this audience, here we just summarise the main principles of Temporal Abstraction.

2.1. Temporal Abstraction

Temporal Abstraction has become a hot topic in Medical Informatics in the recent years. The main principles have been by outlined by Shahar [Sh97]. The idea of Temporal Abstraction is to describe a temporal sequence of values, actions or interactions in a more abstract form, which provides a tendency about the status of a patient. For example, for monitoring the kidney function it is fine to provide a daily report of multiple kidney function parameters. However, information about the development of the kidney function on time and if appropriate even an early warning against a forthcoming kidney failure means a huge improvement [SPG99].

To describe tendencies, an often-realised idea is to use different trend descriptions for different periods of time, e.g. short-term or long-term trend descriptions etc. [e.g. Mi95]. The lengths of each trend description can be fixed or they may depend on concrete values (e.g. successive equivalent states may be concatenated).
However, concrete definitions of the trend descriptions depend on characteristics of the application domain:

(1) On the number of states and on their hierarchy,
(2) On the lengths of the considered courses, and
(3) On what has to be detected, e.g. long-term developments or short-term changes.

3. Prognostic Model for TeCoMed

Since we believe that warnings can be appropriate in about four weeks in advance, we consider courses that consist of four weekly incidences. However, so far this is just an assumption that might be changed in the future. Figure 2 shows the prognostic model for TeCoMed. It consists of four steps (the grey boxes on the right side of figure 2).

3.1. Temporal Abstraction

We have defined three trends concerning the changes on time from last week to this week, from last but one week to this week and from the last but two week to this week. The assessments for these three trends are "enormous decrease", "sharp decrease", "decrease", "steady", "increase", "sharp increase", and "enormous increase". They are based on the percentage of change. For example, the third, the long-term trend is assessed as "enormous increase" if the incidences are at least 50% higher than those three weeks ago. If they are only at least 30% higher it is assessed as "sharp increase", and if they are only at least 5% higher it just an "increase".

Together with the four weekly data these assessments are used to determine similarities between a query course and all courses stored in the case base. Our intention for using these two sorts of parameters is to ensure that a query course and an appropriate similar course are on the same level (similar weekly data) and that they have similar changes on time (similar assessments).

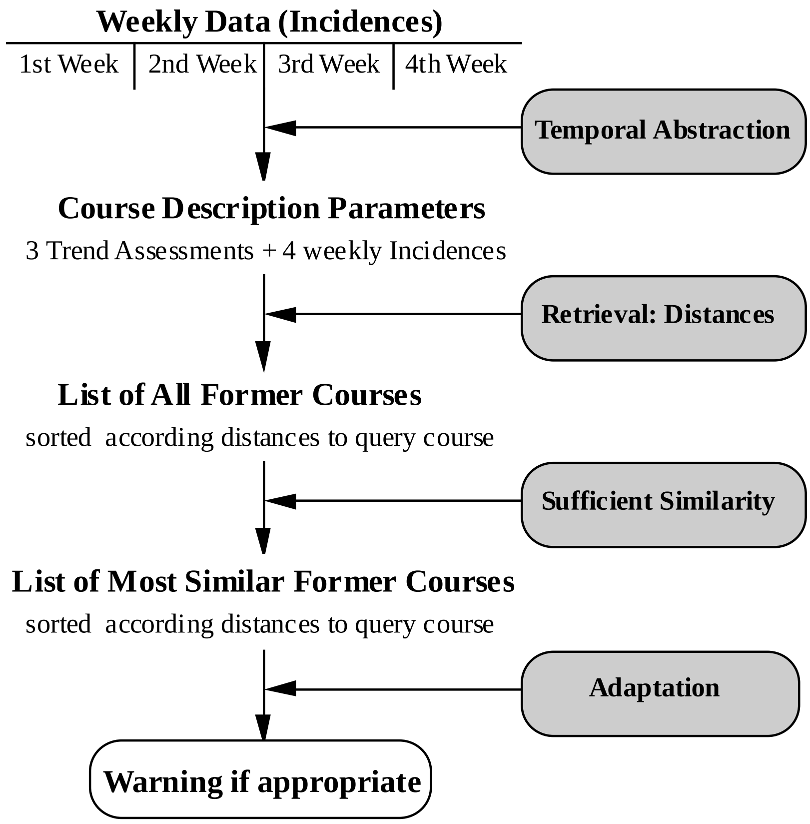

Figure 2. The prognostic model for TeCoMed

3.2. Searching for similar courses

So far, we compute distances between a query course and all courses stored in the case base sequentially. In the future we hope to develop a more efficient retrieval algorithm. The considered attributes are the three nominal valued trend assessments and the four weekly incidences. The similarity measure is only based on our considerations, because no knowledge about it is available and learning by comparing with desired results is already necessary for later step (explained in section 3.5.).

When comparing a current course with a former one, distances between equal assessments are valued as 0.0, between neighbouring ones as 0.5, and otherwise as 1.0 (e.g. "increase" and "sharp increase" are neighbouring). Additionally we use weights; the values for the short-term trend are weighted with 2.0, those for the medium-term trend with 1.5, and those for the long-term trend with 1.0. The idea is that we believe that more recent developments should be more important than earlier ones.

For the weekly data, we compute differences between the values of the query and those of each former course. We divide a difference by the value of the query course and weight it with the number of the week within the four weeks course (e.g. the first week gets the weight 1.0, the current week gets 4.0).

Finally, the distance concerning the trend assessments and the distance concerning the incidences are added.

3.3. Sufficient Similarity Check

The result of computing distances is a very long list of all former four weeks courses sorted according to their distances. For the decision whether a warning is appropriate, this list is not really helpful, because most of the former courses are rather dissimilar to the query course. So, the next step means to find the most similar ones. One idea might be to use a fixed number, e.g. the first two or three courses in the sorted list. Unfortunately, this has two disadvantages. First, even the most similar former course might not be similar enough, and secondly vice versa the fourth, fifth etc. course might be nearly as similar as the first one.

So, we decided to filter the most similar cases by applying sufficient similarity conditions. So far, we use just two thresholds. First, the difference concerning the three trend assessments between the query course and a most similar course has to be below a threshold X. This condition guarantees similar changes on time. And secondly the difference concerning the incidences of the current week must be below a threshold Y. This second condition guarantees an equal current level. Of course further conditions concerning the incidences of the three weeks ago might also be used.

3.4. Adaptation

So, now we have got a usually very small list that contains only the most similar former courses. However, the question arises how these courses can help to decide whether early warning is appropriate. In Case-based Reasoning, the retrieval usually provides just the most similar case whose solution has to be adapted to fit for the query course. As in Compositional Adaptation [WSC98] we take the solutions of a couple of similar cases into account.

The question is: what are the solutions of courses of incidences? The obvious idea is to treat the course continuation of a four weeks course as its solution. However, in contrast to Viboud et al. [Vi01] we do not intend to predict future incidences, but to provide interested people (practitioners, pharmacists etc.) with warnings against approaching influenza waves. So, we have marked those time points of the former courses where we in retrospect believed a warning would have been appropriate; e.g. in the 17^{th} week of the 2000/2001 season (see fig.1). This means that a solution of a four weeks course is a binary mark, either a warning was appropriate or not.

For the decision to warn, we split the list of the most similar courses in two lists. One list contains courses where a warning was appropriate; the second list gets the other ones. For both of these new lists we compute the sum of the reciprocal distances of their courses to get sums of similarities. Subsequently, the decision about the appropriateness to warn depends on the question, which of these two sums is bigger.

3.5. Learning

In section 3.3. we have introduced two threshold parameters X and Y. However, we have not explained how we are getting good settings for them. In fact there is no chance to know them. Since they are very important for the solution, namely the decision whether to warn, we attempt to learn them. For each of our complete influenza courses (from October to March), we have made the same experiment; we used it as query course and we have tried to compute the desired warnings with the remaining courses as case base. Therefor we have varied the values for the threshold parameters X and Y. So far, we have not learnt single optimal values but intervals for the threshold parameters. With combinations of values within these intervals all desired warnings could be computed.

4. First Results and Future Work

So far, we have developed a program that gives early warnings of approaching influenza waves for the German federal state Mecklenburg-Western Pomerania. Since we receive data since 1997, our case base just contains four influenza periods. For each of them, our program is able to compute the desired warnings by using the other three periods as case base. However, the last influenza epidemic in Western Europe, where doctors even ran out of vaccine, occurred in winter 1995-96 [Ca98].

Unfortunately, we do not have data for this period. Nevertheless, we hope to be able to predict such epidemics with the help of our data of the recent influenza waves.
At present the computed warnings and follow-up warnings are only displayed on a machine. In the near future we intend to send them by email to interested people.

Furthermore, so far we have focussed on the temporal aspect of influenza waves for the whole federal state. Very recently, we have started to apply our program to smaller units, namely to 6 cities or towns and to 12 districts in Mecklenburg-Western Pomerania. Since we only receive data of written unfitness for work from the main health insurance company, incidences for some of these units are rather small (even the peaks are sometimes below 100 per week). For such units our program has difficulties to determine whether an increase is already the beginning of an influenza wave or if it occurred just accidentally. The general problem is that the smaller the incidences are, the higher is the influence of coincidence.

References

[Ca98] Carrat, F. et al.: Surveillance of influenza-like illness in France: The example of the 1995-96 epidemic. Journal of Epidemiology Community Health 52 (Suppl1), 1998; S.32-38

[Do83] Dowdle, W.R.: Informed Consent Nelson-Hall, Inc. Chicago, III.

[FB97] Farrington, C.P.; Beale, A.D.: The Detection of Outbreaks of Infectious Disease. In (Gierl, L. et al. Hrsg): First International Workshop on Geomedical Systems, Teubner-Verlag, Stuttgart Leipzig, 1997; S. 97-117.

[Hw99] Hwang, M.Y.: Do you have the flu? In: JAMA 281, 1999; S. 962.

[Le98] Lenz, M. et al. (Hrsg.): Case-Based Reasoning Technology, Springer-Verlag, Berlin Heidelberg New York, Lecture Notes in Artificial Intelligence 1400, 1998.

[Lo69] Lorenz , E.N.: Atmospheric predictability as revealed by naturally occuring analogies. Journal of Atmospherice Sciene 1969; S.26.

[Mi95] Miksch, S. et al: Therapy planning using qualitative trend descriptions. In (Barahona, P; Stefanelli, M; Wyatt, J. Hrsg.): 5th Conference on Artificial Intelligence in Medicine, Springer-Verlag, Berlin Heidelberg New York, Lecture Notes in Artificial Intelligence 934, 1995; S. 197-208.

[Ni95] Nichol, K.L. et al: The effectiveness of Vaccination against Influenza in Adults. New England Journal of Medicine 333, 1995; S. 889-893.

[Pr01] Prou, M. et al.: Exploratory Temporal-Spatial Analysis of Influenza Epidemics in France. In (Flahault, A. et al. Hrsg.): Abstracts of Third International Workshop on Geography and Medicine, Paris, 2001; S. 17.

[Sc96] Schmidt, R. et al.: Prognoses of Multiparametric Time Course Abstractions in a Case-Based Reasoning System. In (Burkhard, H.-D.; Lenz, M. Hrsg): 4th German Workshop on Case-Based Reasoning. Humboldt-Universität Berlin, 1996; S.170-177.

[Sh97] Shahar, Y.: A Framework for Knowledge-Based Temporal Abstraction. Artificial Intelligence 90, 1997; S. 79-133.

[Sh01] Shindo, N. et al.: Distribution of the Influenza Warning Map by Internet. In: (Flahault, A. et al. Hrsg.): Abstracts of Third International Workshop on Geography and Medicine, Paris, 2001; S. 16.

[SPG99] Schmidt, R.; Pollwein, B.; Gierl, L.: Medical multiparametric time course prognoses applied to kidney function assessments. International Journal in Medical Informatics 53 (2-3), 1999; S. 253-264.

[Vi01] Viboud, C. et al: Forecasting the spatio-temporal spread of influenza epidemics by the method of analogues. In: Abstracts of the 22nd Annual Conference of the International Society of Clinical Biostatistics, Stockholm, August 20-24, 2001; S. 71.

[WSC98] Wilke, W.; Smyth, B.; Cunningham, P.: Using Configuration Techniques for Adaptation. In: [Le98]; S. 139-168.