Intelligent Systems
Already Influence Our Lives

Gams, Matjaž; Krivec, Jana

Abstract*— *In this paper we present our recent research in the area of intelligent systems dealing with real-life problems and how that affects actual life worldwide and in Slovenia. Three applications are discussed: the influence of high knowledge on GDP worldwide, predictions of oil prizes and predictions of demographic trends in Slovenia with detailed demographic analysis. It was concluded that a) high knowledge is of major importance for a country's wealth; b) oil prices are not going to be low ever again and c) Slovenia is fast moving towards a demographic suicide which is common thing in most liberal, rich western countries. All these subjects were also presented in media thus hopefully having some impact on the future.

Index Terms— ***artificial intelligence, intelligent systems***

1. Introduction

IN comparison with classical AI academic orientation, the term Intelligent Systems denotes a similar area in a more general, technically and practical oriented way, using AI methods [2,3], classical computational approaches and some cognitive ideas [6], thus adding to computational brute force some "human" intelligence. Ubiquitous computing technologies, systems and solutions are influencing our everyday world and their impact on society is becoming increasingly prevalent. Intelligent systems were first developed for, and often by, industry leaders in the following sectors: manufacturing, mining, forestry, energy production and agri-food [2].

Machine learning and lately data mining are among the most successful application areas of artificial intelligence and intelligent systems [2]. Whenever there are lots of learning examples, these systems learn properties of domain and make predictions about future cases. The constructed knowledge is often in the form of understandable trees, rules and other readable representations. An example of a successful scientific and engineering SW tool are Weka [15] and Orange [4]. These systems are a collection of machine learning algorithms/programs for solving real-world data mining problems. They both contains tools for data pre-processing, modeling and data exploration techniques with learning algorithms and evaluation methods (classification, regression, clustering, association rules, visualization). Last but not least they are very powerful data analytical tools. Especially Orange is very convenient also for advanced research, since it allows users to implement their own analysis methods or even use an existing algorithm and replace some of its standard components with their own ones since it is an open source system / platform.

In the rest of the paper we describe three relevant practical applications demonstrating the power of intelligent systems for real life.

2. High Knowledge and GDP

In our first practical case we used methods of intelligent systems to design an accurate decision tree showing relevance of science and education for country success. The dataset with learning examples was extracted from various statistical databases, provided by The World Bank, UNESCO Institute for Statistics, USAID (Global Educational Database) and WIPO Patent Scope. It consist of 158 examples (countries) with 50 indicators from the year 2001 of which 7 represent economic indicators (e.g. GDP per capita, GDP growth, GNI per capita…), 14 R&D indicators (e.g. Researchers per million inhabitants, GERD per capita, Grants of patents…), 24 educational indicators (e.g. Tertiary students per 100,000 inhabitants, Public expenditure on education as percentage of GDP, School life approximation…) and 5 general indicators (e.g. Fixed line and mobile phone subscribers per 1000 people, Internet users per 1000 people, Military expenditure as percentage of GDP). All indicators are numeric except one discrete. Discrete indicator Gross National Income (GNI) per capita was chosen for the class. GNI prizes the total value of goods and services produced within a country (i.e. its Gross Domestic Product) together with its income received from other countries (notably interest and dividends) and less similar payments made to other countries. The indicator can take one of the three values, i.e. low, middle or high. From the ML and DM techniques available in Weka and Orange we have chosen J48, the implementation of Quinlan`s C4.5 [15], a

Manuscript received February 1, 2007.

M. Gams is with the Jožef Stefan Institute, Department of Intelligent Systems, Ljubljana, Slovenia

J. Krivec is with the Jožef Stefan Institute, Department of Intelligent Systems, Ljubljana, Slovenia (e-mail: jana.krivec@ijs.si)

technique used for the induction of classification trees. We conducted four experiments using different sets of indicators, trying to find as accurate and informative classification tree as possible. In all experiments default values of the classifier parameters set in Weka and Orange were used. We experimented with the usage of reduced error pruning, but the obtained classification trees were less accurate. To estimate the accuracy of the trees, we used 10-fold cross-validation.

One of the most interesting and accurate (70.25%) trees we've got is shown in Figure 1. In this case we have concentrated on higher education and R&D, and chosen 26 relevant indicators. It can be seen from the tree (Figure 1) that the high income countries invest more in R&D, have advanced tertiary education, more granted patents and higher percentage of high-technology exports. On this base we can conclude that high knowledge plays one of the most important roles in economic welfare. As usual in case of classification trees, we can use it also for predicting the income group of some new country on the basis of indicators of education and R&D. Some exceptions were observed, e.g. countries rich in oil that were much richer than compared to the above criteria. Detailed analysis is presented in [14].

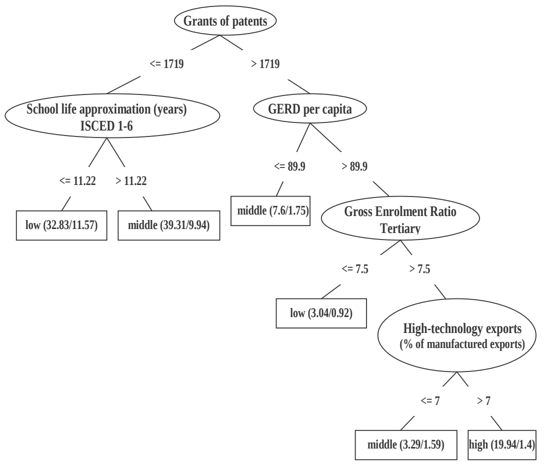

Figure 1: The influence of science and education on GNI per capital.

3. Oil Prices

Intelligent systems are of major importance also when prediction of future events is concerned. They are able to find out typical patterns of events in the past and use them to make projections for the near future.

In September 2004, the first author of this paper presented a popular paper in the Delo daily [7]. With the help of various separate methods of intelligent systems, combing them with some basic background knowledge and constraints such as supply and reserves of cheap and over-all oil, it was soon becoming clear that new developing countries/economies like China are already increasing consumption/demand to a level where production can not generate extra reserves, thus oil prices will not ever be cheap as they used to be for most of the last 50 years (see Figure 2).

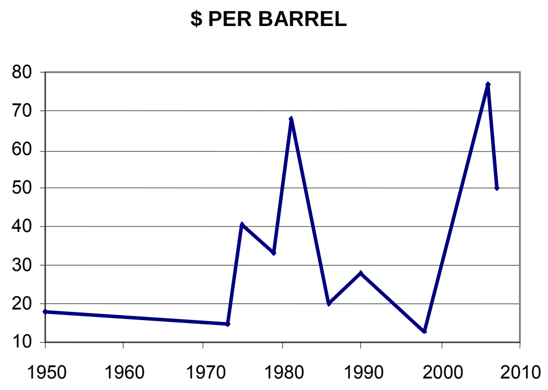

Figure 2: Oil prices

At that time, at least half of the oil experts, among which L. Raymond and J. Browne, both CEO`s of major oil companies, predicted differently. They claimed that the oil prices will eventually decrease, even more, that they will return to 35$ per barrel. At the same time, most of the experts predicted that the peak of oil production will be reached in around two decades.

The prices of oil in 2007 were around 60-100$ per barrel and there is no real indication that they will drop significantly with the exception of normal fluctuations. Essential for the prediction was to realistically estimate relations through various methods and make a common-sense integration. Simple predictions without taking in concern this essential background knowledge have proven to be useless.

While alternative fuels will and are popping at costs around 70$ per barrel, the overall effect of the rising new underdeveloped countries is bringing havoc to the developed world and standard. It is estimated that most of the population in the developed world now lives worse than 5 years ago in terms of energy and food (not in terms of informatics and communications since these areas are not directly related to prices of oil.)

4. Slovenian Demography

In another application we analyzed Slovenian demographic trends. First, we predicted the number of citizens in Slovenia (see Figures 3 and 4) in the next 50 years. Again, there was some background knowledge such as a generally accepted equation for citizen growth in a period of time (a year for example):

Growth = births + incomings – deaths – outcomings

where incomings and outcomings are known as migration. A typical property of this kind of equations is that predictions tend to be exponentially growing or declining.

The basic dataset was collected from Statistical office of the Republica Slovenia (Stat.Base). Three key parameters were concerned: fertility, migrations and life expectancy incensement. They were first set to project overall population in Slovenia, giving the same result as UN and Eurostat predictions (see Figure 3) [5,12,13].

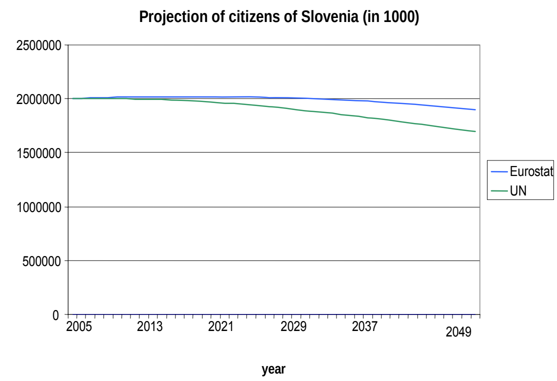

Figure 3: Prediction of number of citizens in Slovenia

Overall, these predictions show that the number of citizens of Slovenia will not fade drastically by the year of 2050. However, we concluded that increased migrations drastically affected the final result. Furthermore we wanted to find out what will happen with the native Slovenians in the future.

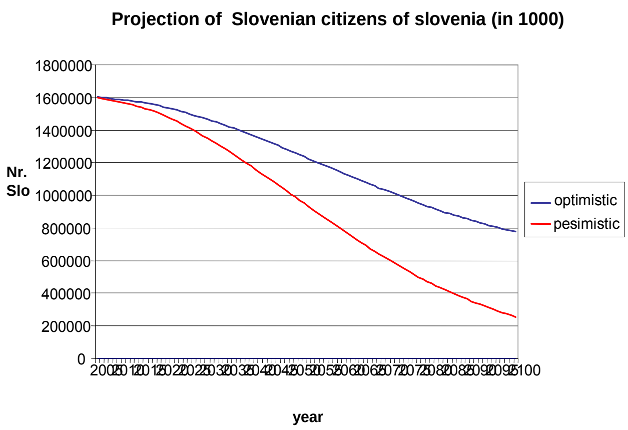

Figure 4: Prediction of number of Slovenian citizens in Slovenia

With this object to investigate we took data structure of native Slovenians in the year 2002 (census of population), and projected them on the year 2100. The first graph in Figure 4 is adopted to the Eurostat model with the next corrections: fertility 1.2 (as in the year 2005) growing to 1.5 (in the year 2050), constant (after 2050); life expectancy is extending 1 year per 7 years and migration to be 0. This is meant to be a rather optimistic evaluation. The other settings of these parameters is reasonably pessimistic: assuming that migration is -1000, fertility 1.1 and no life expectancy extensions. Both predictions indicate serious problems for economy and quality of life.

4.1. Demographic Analysis of the Causes

Our next attempt was to find out what are the reasons for differences in fertility ratio in different world countries. The analysis was made for 145 countries, on 95 attributes, which we thought might influence the fertility rate of a country. Attributes could be consolidated in 4 groups: country politic/society (39 attributes like country attitude toward maternity leave, heterosexuality, religion, abortion, military,…), economical status (12 attributes like unemployment rate, GDP,…), education and R&D (28) and biological factors (6, for example number of citizens, number of men on 1000 women, etc.). Data were extracted from Wikipedia, UN and Eurostat sources [5,13,16,17]. For the class we have chosen an average number of children that a woman had in her fertility period, assuming that she has more than 49 years [11]. The class indicator can take one of the two values: increase (if the rate is two or more) or decrease (if the rate was less than two).

We analyzed influences of various attributes using machine learning techniques mentioned in the introduction (J48), experimented with selected number of countries and attributes. To estimate the accuracy of the trees, we used 10-fold cross-validation as common.

One of the interesting trees is presented in Figure 5. In this case we took in consideration all the countries and extracted only 11 attributes determining social attitudes towards common life questions like homosexuality status, number of suicides on 1000 people, abortion status, divorces frequency, percentage of women in the parliament, etc.. Our main finding in this case was that fertility rate is higher in more conservative countries (that don't allow abortions, homosexual adoptions etc.) and lower in more liberalistic countries.

Since the countries considered as liberalistic are mostly economically developed, one should be cautious with quick conclusions regarding social relevance. But economical factors proved out not to be among the most important indicators (for further details see [23]).

Additionally we extracted main factors that influence fertility rate within the countries with high economical growth (1000$ per habitant among year 2002 and 2007). 39 such countries appeared. One of the most controversial and accurate (89,74) tree that emerged in this case is presented in Figure 6.

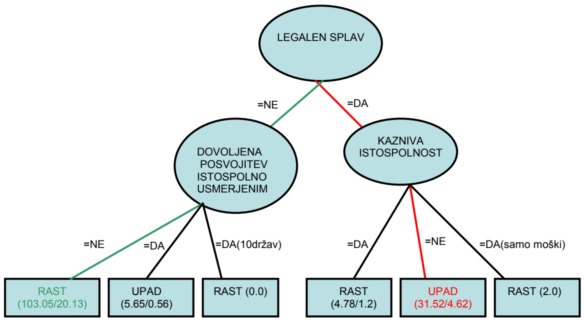

Figure 5. General social attitude towards common life questions.

Does this mean that we can claim with 89,74 accuracy that the length of maternity leave is in correlation with decreased fertility rate? Doubtfully, since other factor like economical status influence the results and furthermore, the trees do not indicate the direction of relation. This shows us that we have to be very careful with jumping to conclusions from ML results. We have to interpret them rationally, with safe amount of caution. Still, there is a hypothesis that lengthening a maternity leave is not the most efficient way of increasing fertility rate.

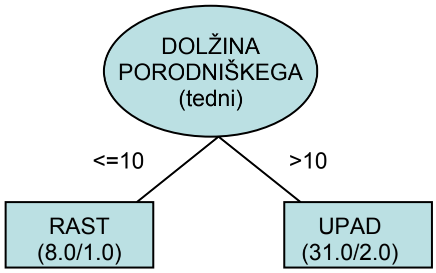

Figure 6. Influence of the maternity leave length on fertility rate.

Our general analysis confirmed former expert predictions and interpretations that fertility rate to a great length depends on economical status of the country. However, more detailed analyses pointed out some even more important factors like general social opinion about personal values and health insurance. Some of the discovered relations are original and rather easy to implement. At the same time it should be noted that the demographic problem is of a long-term nature and that it can not be solved either quickly or easily.

With some publications of the results in mass media [8] we attracted attention of our society and politicians.

5. Conclusion

The importance of intelligent systems steadily grows in recent decades. We have shown the effectiveness of intelligent systems for a couple of relevant questions in Slovenia and world-wide. In particular we have shown that:

- high knowledge is of major importance for a country's wealth, thus silencing critics claiming that less funds should be devoted to science and high education,
- Slovenia should be at least aware that the demographic trends show not only national, but also commercial dark future, so major attention and an action plan are highly desired,
- some of the demographic actions seem pretty cheap and efficient yet introduce major political dilemmas: conservative countries are much successful in terms of fertility rate while rich liberal countries are facing a demographic suicide.

References

[1] Achieving Human-Level Intelligence through Integrated Systems and Research, AI Magazine, vol. 27, no. 2, summer 2006.
[2] AI past, present, and future, AI Magazine, vol. 29, no. 4, pp. 53-97, 2005.
[3] Bratko, I., Muggleton S., "Applications of Inductive Logic Programming," Communications of the ACM, Vol. 38, No. 11, pp. 65-70, 1995.
[4] Demsar J., Zupan B., Leban G., Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana, 2004.
[5] Eurostat http://epp.eurostat.ec.europa.eu/
[6] Gams, M., "Weak intelligence: Through the principle and paradox of multiple knowledge," Advances in computation: Theory and practice, Vol. 6, Nova sc. pub., inc., NY, pp. 245, 2001.
[7] Gams, M., ""Inteligentno" napovedovanje cen nafte : umetna inteligenca in inteligentni sistemi," Delo (Ljubl.), 2.9.2004, str. 18, Znanost.
[8] Gams, M., "Ekonomski in demografski samomor Slovenije," Finance, 9.03.2007, št. 47, str. 12.
[9] Gams, M., "Osnovna demografska gibanja", In Proceedings of the 10th International Multiconference IS 2007, vol. B, Ljubljana, Slovenia,pp.24-27.
[10] Gams, M., Krivec, J., "Analiza vplivov na rodnost", In Proceedings of the 10th International Multiconference IS 2007, vol. B, Ljubljana, slovenia,2007, pp.18-23.
[11] Malačič, J., Demografija: teorija, analiza, metode in modeli, (EF, Učbenik). 6. izd. Ljubljana: Ekonomska fakulteta, 2006. VIII, ISBN 961-240-080-6. [COBISS.SI-ID 224724480]
[12] Stat.Baza Statistični urad Slovenije http://www.stat.si/
[13] UN http://esa.un.org/unpp/
[14] Vidulin, V., Gams, M., "Vpliv investicij v izobraževanje in R&R na gospodarsko rast," Elektroteh. vestn., 2006, vol. 73, št. 5, str. 285-290.
[15] Witten, I. H., Frank E., "Data Mining – Practical Machine Learning Tools and Techniques (sec. ed.)," Morgan Kaufmann, 2005.
[16] WIKI-O http://en.wikipedia.org/wiki/Overpopulation
[17] WIKI1 http://en.wikipedia.org/wiki/Population_growth