

Integrating DROOLS and R software for intelligent map system

Jan Ruzicka

Institute of Geoinformatics

VSB – TU of Ostrava

jan.ruzicka vsb.cz

Keywords: expert system, map sheet evaluation, DROOLS, R software, ontology

Abstract

The paper describes intelligent map system that allows to check errors in map sheets or to help with a map sheet creation. The system is based on expert system DROOLS, ontology created in Protégé and statistical software R. Prototype of the system should evaluate that this kind of integration is possible, so the system is not full of rules. The prototype is filled with twenty rules written in DRL language and with more than thirty items from the ontology. The paper should show how all of these components can be integrated together to allow such kind of a map sheet evaluation. The system is now used for selection of the best method for data classification. The selection is suggested by DROOLS system that uses R software to perform statistical tests of normality and uniformity.

Introduction

The world of cartography is changing, we can see it in any map that is available on the web these days. Any internet user can create own map without any basic knowledge about cartography. There are tools for a map creation available free of charge and geodata available free of charge as well. When the tool for a map creation keeps the process of a map creation under its supervision, the resulting map is usually correct in a term of cartography rules. When the tool gives a lot of options how to create the map, the map is usually full of mistakes. This is described in [1]: "Process of making map is core of the whole cartography, but not only specialists are making maps nowadays. In last years, this process not involved only the cartographers, but also the common users. Production of map with using adequate software is a simple process now, which is used by non-cartographic users. These users do not know basic cartographic rules for making maps and they make maps intuitively. This situation needs the implementation of principles of cartography directly into the map production systems in pursuit of correct and effective maps producing. Instead of the final map is also important the explanation and the proposing of several possible solutions. According to progress in the artificial intelligence, the knowledge-base systems can be applied for this problem. These systems can partly substitute a role of the expert in this process." We have decided to research possibilities how to create intelligent map system, that can help in a process of a map creation. Several tools has been inspected and tested for purposes of the system development. We have discovered that such system can not be simply created with one tool, but that several independent systems should be integrated together. This article describes integration of expert system and statistical system.

Aim of the system

Aim of the system is to help with a map creation for users that are not familiar with cartographic rules. The system can help in two ways:

- answer a question in a process of a map sheet creation,
- check a created map sheet for mistakes.

When the user creates a map there are always steps where he/she must do a decision. For example which size of a font to use for a title of the map or which classification method to use for creating classes breaks. The user just simply (or sometimes not so simply) answers to system questions and obtains recommendations how to finish the step of the map creation. The answers can be filled in a simple graphical user environment with items such as text field or check box. Several answers can be derived from the data used by user for the map creation. A similar approach has been used in the Descartes project [2] and we just adopted it to our project.

Another way, not yet researched in depth in any founded article, is based on checking existing (created) map for mistakes. In this approach the system obtains the map from the user and analyses its content. When it is needed, the system can ask the user for original data. The map is checked according to cartographic rules. This approach is mentioned in [3], but the system mentioned in the paper was not tested and not even developed. A result of the check of the map for mistakes can be of three types:

- a list of mistakes and suggestions how to avoid them,
- a map without mistakes based on the original map,
- a map without mistakes based on the original data and original map.

The simplest way is to provide the user with a list of mistakes and some suggestions how to avoid them. We can generally declare that system described in this article works according to this simplest way. The more difficult is to repair the map. To be able to repair the map there must be met several conditions:

- the map must be available in the form of file, that uses structures that can help with simple map repair (e.g. Scalable Vector Graphics format),
- the mistakes must be from the selected types, not all mistakes can be automatically repaired,
- the original data must be available.

The aim of the system is not to create it so flexible that it is able to find any mistake in a map, but it should be able to find several most horrible mistakes to help with a map quality improvement. For example to avoid creation of maps such as on the following figure (Figure 1).

Pilot project focus

The pilot project is focused only on selected part of cartography techniques namely Choropleth maps and Cartograms. It has been tested on Atlas of Fire Protection in the Czech Republic

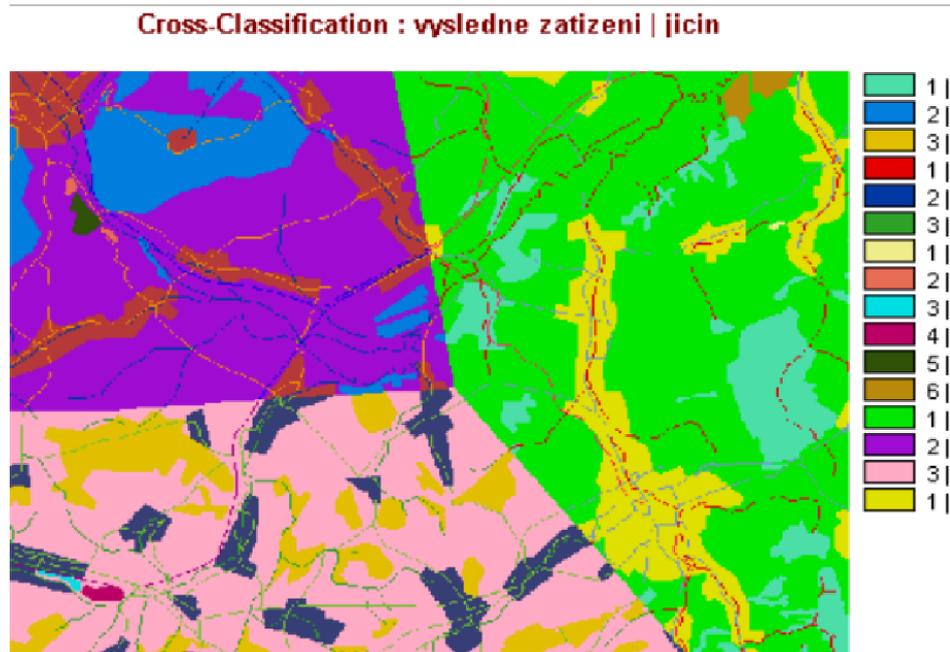


Figure 1: A map with several mistakes

(Ministry of Interior).

The atlas allows to create a choropleth map or a cartogram based on statistical database of events that required fire brigade action. The atlas allows to specify following conditions:

- year from/to of events,
- type of events (e.g. fire where were injured fireman),
- statistical method for generating class intervals (Jenks, Equal interval, etc.),
- number of classes,
- type of frequency (square km, population),
- start colour, end colour for classes visualization.

The user must specify these conditions. Selection of the statistical method is in the pilot project now based on the intelligent map system.

The resulting map can be as on the following figure (Figure 2).

System architecture

The system is based on integration of several items listed on the following figure (Figure 3).

The process of answering to the question which classification method to use is covered by following steps:

- Client (Any SOAP/REST capable – in our pilot project the client is the Atlas) sends data for classification to service.

PEOPLE RESCUED FROM FIRES IN THE CZECH REPUBLIC DISTRICTS AND REGIONS
(January 1, 1997 - December 31, 2005)

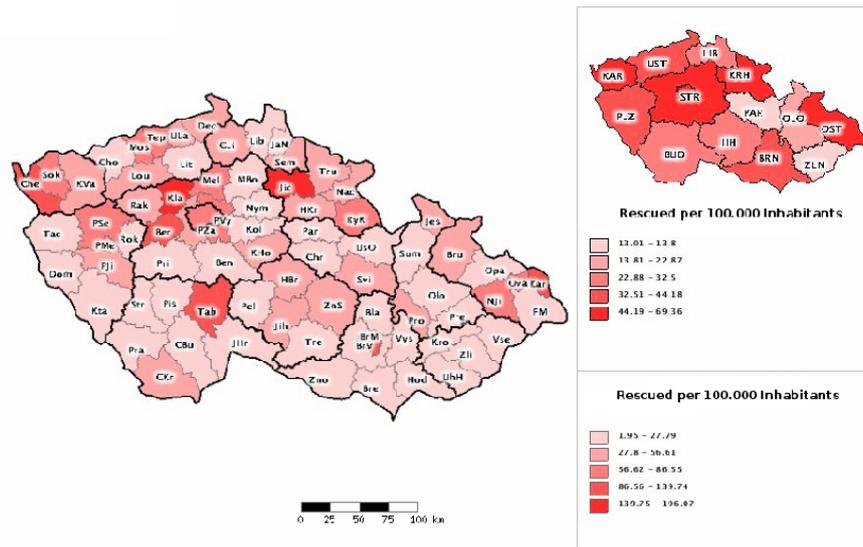


Figure 2: Choropleth map from the Atlas

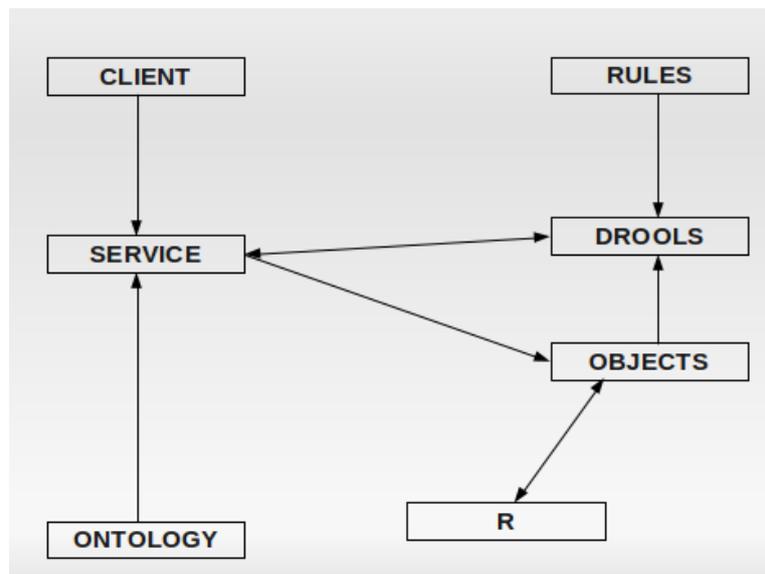


Figure 3: System architecture

- The service reads an ontology (available in OWL format) and creates objects that will be placed in a session of an expert system based on DROOLS.
- When is created an instance of a class named StatisticalValuesGeo, the data from the client are stored into the instance.
- After the data are stored in the instance the instance creates R software instance and runs tests of the data in the R software instance.
- The service creates the session of the expert system and fires all rules on the session.

- Results of the all rules run is stored in the InfoContainer class.
- The service reads results from the InfoContainer class and returns response containing the results to the client.

Ontology

The used ontology is created in Protégé software. The ontology is created with regard to limits of export to Java classes. The export is done via Protégé-OWL-API that has several limits when exporting ontology. So the ontology is just a simple hierarchy with super-classes and sub-classes. The classes have defined attributes with a data type definition and a cardinality relationship between class and attribute.

Class StatisticalValuesGeo

The class StatisticalValuesGeo extends class StatisticalValues, that is defined in the ontology. The extension is based on reaction to the process when data used for a map are stored within this class. In that moment is tested their statistical distribution.

The distribution is tested only for three possible models:

- Normal distribution.
- Uniform distribution.
- Other distribution.

The test of distributions is done in software R via tool rJava (JRI – <http://rosuda.org/rJava/>). The tool rJava is a Java native interface to R software.

Normal distribution

The normal distribution is tested with Shapiro test (module shapiro.test). When the value of resulting W is more than 0.95 and value of resulting pvalue is more than 0.05 then the data are identified as they have normal distribution. See the following code for details.

```
private static boolean testNormality(Rengine re) {
    long e=re.rniParse("shapiro.test(p)", 1);
    long r=re.rniEval(e, 0);

    REXP x=new REXP(re, r);
    RVector rv =x.asVector();

    x = rv.at(0);
    double W = x.asDouble();
    x = rv.at(1);
    double pvalue = x.asDouble();

    if (W > 0.95 && pvalue > 0.05) {
        return true;
    } else {
        return false;
    }
}
```

Uniform distribution

The uniform distribution is tested with Kolmogorov – Smirnov test (module `ks.test`). For the purposes of the test is created uniform distribution based on minimum and maximum from the tested data distribution. When the value of resulting D is less than 0.1 and value of resulting p value is more than 0.05 then the data are identified as they have uniform distribution. See the following code for details.

```
private static boolean testUniformity(Rengine re) {
    REXP x = re.eval("y=c(min(p):max(p))");
    long e=re.rniParse("ks.test(y, p)", 1);
    long r=re.rniEval(e, 0);
    x=new REXP(re, r);

    RVector rv =x.asVector();
    x = rv.at(0);
    double D = x.asDouble();
    x = rv.at(1);
    double pvalue = x.asDouble();

    if (D < 0.1 && pvalue > 0.05) {
        return true;
    } else {
        return false;
    }
}
```

DROOLS

The DROOLS system is used for testing defined cartographic rules. The rules are written in DRL language. For example of decision which classification method to use are used following three rules.

```
rule "NormalDistribution"
when
    N_StatisticalValuesGeo ( distribution == "Normal" )
then
    InfoContainer.method = "EqualInterval";
end

rule "UniformDistribution"
when
    N_StatisticalValuesGeo ( distribution == "Uniform" )
then
    InfoContainer.method = "Quantile";
end

rule "OtherDistribution"
when
    N_StatisticalValuesGeo ( distribution == "Other" ) or \
    N_StatisticalValuesGeo ( distribution == "Unknown" )
then
    InfoContainer.method = "Natural";
end
```

There are also rules to test when is used another classification method that is recommended by the system according to detected distribution.

```
rule "NormalDistributionClassificationScheme"
when
    not EqualIntervalScheme() and N_StatisticalValuesGeo ( distribution == "Normal" )
then
```

```
        InfoContainer.sendMessage("[10] When is distribution of the data normal, the Equal Interval \
classification scheme should be used");
end

rule "UniformDistributionClassificationScheme"
    when
        not QuantileScheme() and N_StatisticalValuesGeo ( distribution == "Uniform" )
    then
        InfoContainer.sendMessage("[11] When is distribution of the data uniform, the Quantile \
classification scheme should be used");
    end

rule "OtherDistributionClassificationScheme"
    when
        not JenksScheme() and N_StatisticalValuesGeo ( distribution == "Other" )
    then
        InfoContainer.sendMessage("[12] When the distribution of the data is not normal or linear, \
the Jenks classification scheme should be used");
    end
```

Problems

There were several problems with integration of R software with DROOLS system.

Integration to JBOSS

When is system run in a single user environment, then there is not any problem. When we decided to place the prototype into JBOSS application server (that is our primary application server for running DROOLS system in multiple user environment), then the R engine instance does not end in the correct way. There was not still the time to fixed this problem so we use simple workaround. There is used solution based on stateless CGI interface that is called from JBOSS engine.

Where to implement call of R

After several test it is not still clear where to place code for running R software. More general solution could be based on separated class, that will work as a proxy for classes that would like to use capabilities of R software. This solution should be used only when the problem with integration to JBOSS will be fixed. At a moment is all code written in the class StatisticalValuesGeo.

Server with older version of R

The server dedicated for our test purposes uses old version of R software and it is difficult to move to newer version. So we had to handle two problems:

- The older version does not support function assign. We fixed this with simple convert array to vector.
- The R engine must be run with `-vanilla` parameter. This was quite hard to find out, because nobody mentioned this problem on any discussion forum.

Conclusion

As a part of our research we did integration of DROOLS with R software. Our findings are simple, but possibly valuable:

- The integration is possible, but at the moment not with a good performance (CGI workaround)
- The solution based on integration of DROOLS and R software allows in the future to use another functions from R.
- R engine can be replaced with another tool (the solution is not directly dependent on the R engine).

References

1. Brus J., Dobesova Z., Kanok, J.: Utilization of Expert Systems in Thematic Cartography. In: Badr Y., Caballe S., Xhafa F., Abraham A., Gros B. (Ed.), INCOS '09 Proceedings of the 2009 International Conference on Intelligent Networking and Collaborative Systems. pp. 285–289. IEEE Computer Society Washington, DC USA. ISBN 978-0-7695-3858-7 (2009)
2. Andrienko, G., Andrienko, N.: Knowledge engineering for automated map design in DESCARTES. In: Advances in Geographic Information Systems, ed. by Medeiros, C.B., 7th International Symposium ACM GIS'99, Kansas City, November 1999 (ACM Press, New York 1999) pp. 66-72 (1999)
3. Růžička, J.: Pomohou webové služby odstranit noční můru kartografů?. 16. kartografická konference (Mapa v informační společnosti)., Univerzita Obrany, 2005, s. 1-10. ISBN 80-72-310-151 (2005)