Segmenting and smoothing territories in the context of non-life insurance

submitted by Desjardins General Insurance Group

Our non-life insurance company (offering fire, accident, and diverse risks insurance) is currently using statistical modelling techniques in order to quantify the risks insured by the company. These models enable us to segment our clients into distinct risk groups. The higher the risk of a segment, the higher the insurance premium for this segment.

Among the inputs to these models, one finds, among other factors, the characteristics of the client (for instant his age or the number of years of ownership), the characteristics of the insured good (for instance the construction year, the kind of roof, or the past accidents), and characteristics such as the use of the insured good (office or single-family home, for instance) or its geographical location (specified through spatial variables, in particular the census data from Statistics Canada). It is well known that risk varies as a function of the geographical location (for instance in the case of theft): quantifying this relation accurately is of the utmost importance.

Because the Canadian market is very competitive, we need high-performance pricing methods and it is thus necessary to quantify incurred risk in a precise manner. It is also essential to improve these models (for instance by including new data) and to explore new methodologies.

In this context Desjardins GAG wishes to take a new look at the geographical segmentation of risks. For the time being the company uses a combination of GBM (Gradient Boosting Method) and spatial smoothing based on Markov random fields: this method, however, is hard to calibrate and lacks robustness.

The goal of this project is to build one statistical model (or several models) based on the aforementioned variables in order to obtain better estimates of the risks incurred by our clients. For the sake of consistency (with respect to clients), these estimates must verify a spatial-continuity constraint, i.e., the solution comprised of the estimates should be smooth. For instance, the estimates for two clients who are neighbours and share the same characteristics should be similar.

The method we are looking for should be robust and display similar performances for various parameters and data sets. The optimization of hyperparameters should be stable, reproducible, easy to implement, and fast (ideally). We should be able to apply this method in the case where there is little risk exposure in a given territory.