

What is the probability that the sun will rise tomorrow?

Written by Marinka

Thursday, 04 August 2011 16:50 - Last Updated Wednesday, 30 January 2013 23:42

It has been some time since my last post, but here is the new one. Perhaps the title sounds a bit inappropriate, but indeed it is well suited. Read till the end, where I explain it for those not figuring it yet (or consider it a puzzle :))

So, what have I been up to lately? Despite summer holidays I have been involved in quite a few projects.

First, [GSoC](http://code.google.com/soc/) Matrix Factorization Techniques for Data Mining project for [Orange](http://orange.biolab.si/) has been progressing well. Code is almost finished, no major changes in framework, factorization/initialization methods, quality measures, etc. are expected. Project is on schedule and has not diverged from initial plan, all intended techniques (plus a few additional I have found interesting along research) are implemented. I have been doing some testing, and have yet to provide more use cases/examples along with thorough explanation and example data sets. I will not go into details here, as implemented methods' descriptions with paper references are published at [Orange wiki project site](http://orange.biolab.si/trac/wiki/MatrixFactorization). The project is great, a mix of linear algebra, optimization methods, statistics and probability, numerical methods (analysis if you want to read some convergence or derivation proofs) with intensive applications in data mining, machine learning, computer vision, bioinformatics etc. and I have been really enjoying working on it, here is my [post](http://orange.biolab.si/blog/2011/06/24/orange-gsoc-mf-techniques-for-data-mining/) at Orange blog. The Orange and its GSoCers have been spotlighted at [Google Open Source Blog](http://google-opensource.blogspot.com/2011/08/whos-new-in-google-summer-of-code-part.html).

Next, there is some image processing; segmentation, primary and secondary object detection, object tracking, morphology measures, filters etc. (no details).

Minor for keeping contact with MS world, [Sharepoint Server 2010 \(SP 10\)](http://sharepoint.microsoft.com/en-au/Pages/default.aspx). I have some experience with it (and its previous version MOSS 2007), both in administration and especially in code. This time it was not about coding workflows using [Win Workflow Foundation](http://msdn.microsoft.com/en-us/netframework/aa663328), developing Web parts/sites/custom content types/web services (...) but providing an in-site publishing hierarchy for data in custom lists and integration with Excel Services (not with new 365 Cloud service). Obstacles were limited server access (hosting plan), old versions of software and usual MS stuffs (:)). In SP 10 these are SPFieldLookups filters and cascading lookups, data connections between sites/lists/other content. As always there are some nice workarounds which have resolved all issues.

Last (not least) I have been catching up with all the reading material I was forced to put aside during the year (well not entirely true: the more I read, the more should be read, so the pile of papers in iBooks and Mendeley app is not getting any smaller :)).

Here we are, what about the post's title? [The sunrise problem](http://en.wikipedia.org/wiki/Sunrise_problem) was introduced by Laplace (french mathematician known for Bayesian interpretation of probability, Laplace transform, Laplace equation, differential operator, work in mechanics and physics). Is the probability that the sun will rise tomorrow 1 if we can infer from the observed data that it has risen every day on record? :) So what is the answer of the question in the title? The inferred probability depends on the record - whether we take the past experience of one person, humanity, or the Earth history. This is the

What is the probability that the sun will rise tomorrow?

Written by Marinka

Thursday, 04 August 2011 16:50 - Last Updated Wednesday, 30 January 2013 23:42

[reference class problem](http://en.wikipedia.org/wiki/Reference_class_problem) - with Bayes any probability is the conditional probability given what a person knows. Simple principle emerged from this, add-one or Laplacian smoothing (Example: Doing spam email classification with a bag of words model or text classification with multinomial model, this allows the assignment of positive probabilities to words which do not occur in the sample) and corresponds to the expected value of the posterior.