

APPROCHE BAYÉSIENNE HIÉRARCHIQUE ET DYNAMIQUE POUR LA PRÉDICTION ET L'AGRÉGATION DE SCORES

Adrien Todeschini¹, Vincent Couallier², Marie Chavent³

¹*Scorelab.io*, adrien@scorelab.io

²*U. Bordeaux, IMB (UMR 5251)*, vincent.couallier@u-bordeaux.fr

³*U. Bordeaux, IMB (UMR 5251), Inria*, marie.chavent@u-bordeaux.fr

Résumé. Scorelab a développé le Global Wine Score, un agrégateur de scores utilisé pour évaluer et comparer l'ensemble des vins du monde. Il repose sur les notes de critiques reconnus dans le monde du vin. Cependant, tous les vins ne sont pas notés par un nombre suffisant de critiques pour pouvoir délivrer un degré de confiance minimal. Nous développons une approche bayésienne pour enrichir le score d'une prédiction qui emprunte de l'information aux millésimes antérieurs ainsi qu'aux appellations voisines. Nous exploitons la hiérarchie géographique des appellations par le biais d'un modèle bayésien hiérarchique. Les notes résiduelles au cours des millésimes sont modélisées par un filtre de Kalman multi-capteur.

Mots-clés. Agrégation de scores, modèle bayésien hiérarchique, filtre de Kalman multi-capteur

Abstract. Scorelab has developed the Global Wine Score, a score aggregator used to evaluate and compare all the wines in the world. It is based on the ratings of critics who are recognized in the world of wine. However, not all wines are rated by a sufficient number of critics to be able to provide a minimal degree of confidence. We develop a Bayesian approach to enrich the score with a prediction that borrows information from earlier vintages and neighboring appellations. We exploit the geographical hierarchy of appellations through a hierarchical Bayesian model. Residual ratings over the years are modeled by a multi-sensor Kalman filter.

Keywords. Score aggregation, Bayesian hierarchical model, multi-sensor Kalman filter

Session invitée – Retours d'expériences PEPS AMIES

Cet exposé prend place dans la session organisée par AMIES (l'agence maths-entreprises) et présentant des retours d'expérience sur le dispositif PEPS (projets exploratoires premier soutien). L'exposé introductif de Gilles Stoltz et Myriam Maumy-Bertrand détaillera ce dispositif, qui permet de co-financer des travaux de recherche (ou recherche et développement) entre une équipe académique et une entreprise (hors grands groupes). Le présent exposé constitue un exemple de projet ainsi soutenu.

1 Contexte

Ce projet PEPS a duré neuf mois de novembre 2016 à août 2017. Scorelab est une jeune startup de six personnes dont le projet est axé autour du développement et de la mise en œuvre :

- d'algorithmes et outils d'agrégation de notes d'experts et de consommateurs fournissant une évaluation à la fois synthétique et objective de produits.
- d'algorithmes et outils de recommandation personnalisée de produits à destination d'e-commerçants.

Le premier domaine d'application visé est celui du vin. Scorelab collecte et structure une base de données importante d'évaluations de critiques de vins (~200k vins, 1M notes). Elle a développé le site www.globalwinescore.com qui fournit un score unique par vin millésimé à partir de l'ensemble des notes collectées.

Dans ce cadre, les missions du PEPS ont tourné autour de trois axes :

- l'étude et la validation de l'algorithme d'agrégation de notes du Global Wine Score.
- l'exploration d'un nouveau modèle de scoring pour réduire l'incertitude sur les vins ayant reçu peu de notes.
- l'exploration d'algorithmes de machine learning pour la recommandation personnalisée de vins.

En parallèle, nous avons également développé une collaboration scientifique entre Scorelab et un groupe international de chercheurs en machine learning ayant pour objectif le développement d'une application web pour la recommandation personnalisée de contenus (articles, posters, exposés) lors de conférences scientifiques. Nous visons les conférences internationales de premier plan dans le domaine du machine learning : NIPS¹ et ICML².

2 Global Wine Score bayésien

Le Global Wine Score (GWS) est le fruit de neuf mois de recherche et développement par Scorelab antérieurs au projet PEPS. L'algorithme développé en Python répond aux principales problématiques posées par l'agrégation :

- l'hétérogénéité des échelles de notation et de leur granularité : certains critiques utilisent une échelle sur 20, d'autres sur 100, d'autres évaluations de consommateurs sont sur 5 étoiles.

¹<https://nips.cc>

²<https://icml.cc>

- hétérogénéité de la perception de la qualité de chaque évaluateur : certains critiques sont plus sévères que d'autres, ont des préférences pour certains types de vin, *etc.*
- hétérogénéité des ensembles de vins dégustés : certains critiques sont spécialisés dans les vins français, d'autres dans les vins italiens.
- l'évolution temporelle des notes : le vin peut être noté en primeur ou bien après plusieurs années de vieillissement.

Cependant, le Global Wine Score n'est calculé que pour les vins millésimés ayant reçu au moins trois notes de critiques. Scorelab souhaite enrichir le score d'un *a priori* historique lié aux évaluations du vin sur des millésimes antérieurs ainsi que l'évaluation de vins similaires (appellation, millésime).

Dans ce but, nous avons développé un modèle bayésien hiérarchique (Gelman *et al.*, 2014) permettant d'estimer le score moyen ainsi que la variance d'une appellation millésimée. La formulation hiérarchique permet d'emprunter de l'information aux appellations voisines et ainsi régulariser l'estimation des appellations ayant peu de notes. Les scores résiduels associés à chaque vin ou château sont modélisés par un filtre de Kalman multi-capteur (Willner *et al.*, 1976) dont les bruits d'évolution et de mesure sont calibrés par maximum *a posteriori*.

2.1 Modèle bayésien hiérarchique

Soit z_{ijk} la note du critique j attribuée au vin i de millésime k , on suppose la vraisemblance suivante

$$z_{ijk} | \nu_a, \tau_a \sim \mathcal{N}(\nu_a, \tau_a^{-1}), j = 1, \dots, n_{ik}$$

où ν_a et τ_a sont le score moyen et la précision de l'appellation du vin i au millésime k . On suppose de plus le modèle hiérarchique suivant sur les moyennes et précisions d'appellations a que l'on considère comme les nœuds d'une hiérarchie basée sur les terroirs (*e.g.* Margaux < Médoc < Bordeaux < France < Monde). Pour tout nœud a appartenant à l'arbre hiérarchique

$$\begin{aligned} \nu_a | \nu_{p(a)}, \tau_{p(a)} &\sim \mathcal{N}(\nu_{p(a)}, \tau_{p(a)}^{-1}) \\ \tau_a | \alpha_{\ell(a)}, \beta_{\ell(a)} &\sim \Gamma(\alpha_{\ell(a)}, \beta_{\ell(a)}) \end{aligned}$$

où $p(a)$ est le nœud parent et $\ell(a)$ est le niveau du nœud a . On suppose de plus un *a priori* vague sur les hyperparamètres du modèle graphique. Nous avons effectué l'inférence via méthodes MCMC (Gilks, 2005) à l'aide du logiciel JAGS (Plummer *et al.*, 2003).

2.2 Filtre de Kalman multi-capteur

On s'intéresse ensuite au comportement du résidu normalisé de la note z_{ijk} par rapport au niveau moyen de son appellation au millésime k . On note $\tilde{t} = \sqrt{\tau_a}(t - \nu_a)$ le changement de variable où ν_a et τ_a sont le score moyen et la précision de l'appellation du vin i au millésime k .

On suppose le modèle suivant pour tout $k \geq 1$ et $j \in J_{ik}$:

$$\begin{aligned} \tilde{x}_{i1} &\sim \mathcal{N}(\tilde{\mu}_{1|0}, \tilde{\sigma}_{1|0}^2) && \text{(initialisation)} \\ \tilde{x}_{ik}|\tilde{x}_{i,k-1} &\sim \mathcal{N}(\tilde{x}_{i,k-1}, \tilde{\sigma}_{xi}^2), k > 1 && \text{(évolution)} \\ \tilde{z}_{ijk}|\tilde{x}_{ik} &\sim \mathcal{N}(\tilde{x}_{ik}, \tilde{\sigma}_{zi}^2) && \text{(mesure)} \end{aligned}$$

où $\tilde{\mu}_{1|0} = 0$ et $\tilde{\sigma}_{1|0}^2 = 1$ sont des paramètres initiaux, $\tilde{\sigma}_{xi}^2$ et $\tilde{\sigma}_{zi}^2$ sont les variances de bruit d'évolution et de mesure du vin i .

Grâce aux équations de Kalman, cette modélisation permet d'obtenir de manière analytique une prédiction sur le score d'un vin $\mu_{ik} = \mathbb{E}[x_{ik}|z_{ik-1}, \dots, z_{i1}]$, avant que celui-ci n'ait reçu la moindre notation. Le score bayésien final (*a posteriori*) est une moyenne pondérée des notes de critiques observées et de la prédiction dont le poids ω_{ik} est relatif à son degré de confiance (voir figure 1)

$$s_{ik} = \mathbb{E}[x_{ik}|z_{ik}, \dots, z_{i1}] = \frac{1}{\omega_{ik} + n_{ik}} \left(\omega_{ik}\mu_{ik} + \sum_{j \in J_{ik}} z_{ijk} \right)$$

Par ailleurs, grâce à l'ajout d'un *hyperprior* $p(\tilde{\sigma}_{xi}^2, \tilde{\sigma}_{zi}^2)$ et le calcul de la vraisemblance marginale $p(z_{i1}, \dots, z_{ik_{\max}}|\tilde{\sigma}_{xi}^2, \tilde{\sigma}_{zi}^2)$, fourni de manière analytique par les équations de Kalman ; nous pouvons estimer les deux paramètres de bruit d'évolution de mesure par maximum *a posteriori*. Nous avons en outre adopté une approche bayésienne empirique pour définir l'*hyperprior*.

3 Conclusion et perspectives

Notre modélisation bayésienne présente de nombreux avantages. Elle nous permet notamment de capturer l'incertitude des quantités à estimer. Notre modèle est également compatible avec un nombre variable d'observations à chaque temps. Le score *a posteriori* s'interprète facilement comme une moyenne pondérée des observations et de la prédiction.

Le modèle est actuellement séparé en deux parties, un modèle hiérarchique et un modèle dynamique. Il pourrait être intéressant de développer un modèle unique de type filtre de Kalman hiérarchique. Nous pourrions également comparer les performances prédictives de notre modèle avec d'autres approches purement machine learning.

La collaboration entre Scorelab et l'IMB a permis la montée en compétence de Scorelab sur différents sujets cruciaux pour son développement. Scorelab est satisfaite de cette

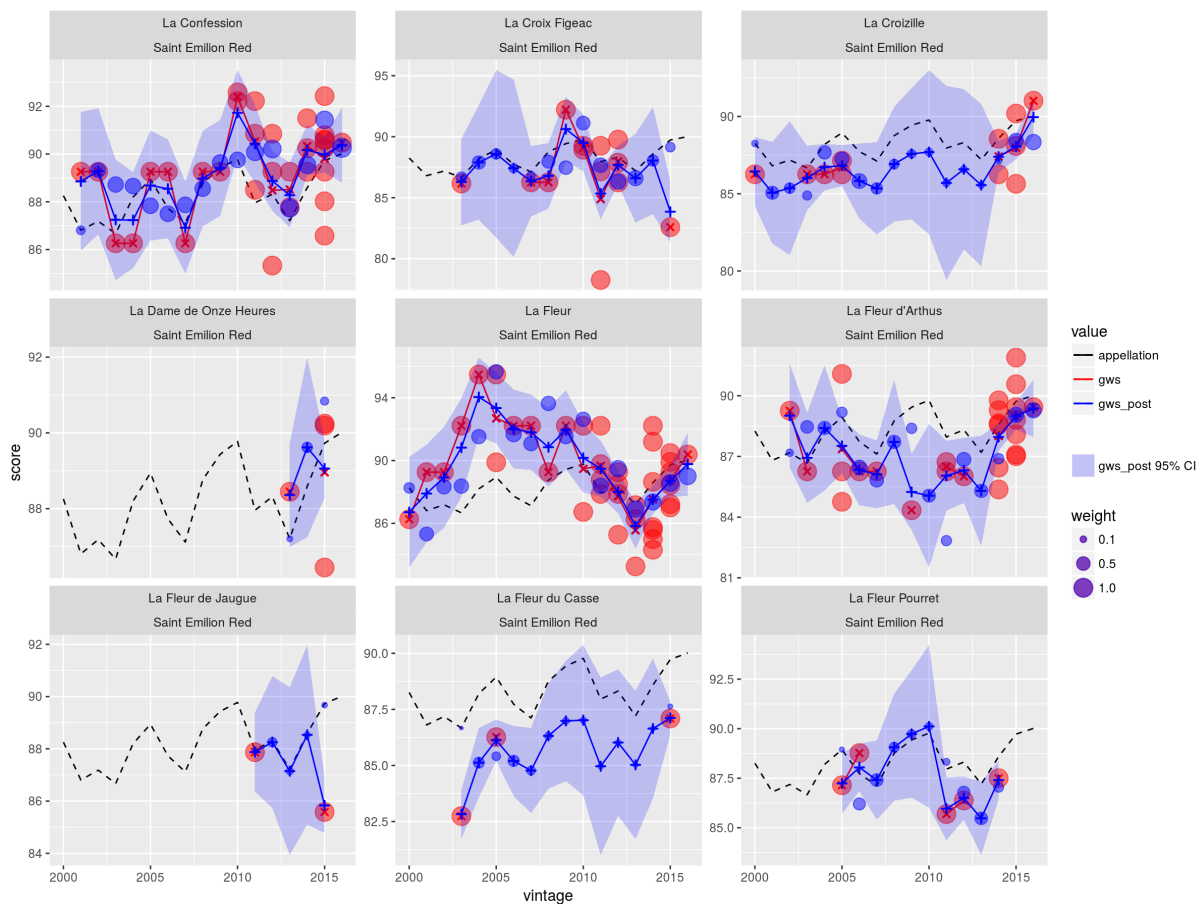


Figure 1: Score bayésien (courbe bleue `gws_post`) et son intervalle de crédibilité en fonction du millésime pour 9 vins. Également présents : les scores prédits (points bleus), les notes de critiques normalisées (points rouges), le score GWS standard (courbe rouge `gws`) et le niveau moyen de l'appellation (courbe noire pointillée).

collaboration et a recruté Adrien Todeschini en CDI et comme associé suite à la collaboration. Scorelab est actuellement dans une phase d'amorçage avec le projet d'une levée de fonds pour mener à bien son programme de recherche et développement sur les prochains 18 mois.

Bibliographie

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. et RUBIN, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.

GILKS, W. R. (2005). *Markov chain Monte Carlo*. Wiley Online Library.

PLUMMER, M. *et al.* (2003). JAGS: A program for analysis of bayesian graphical models using Gibbs sampling. *Dans Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, Austria.

WILLNER, D., CHANG, C. et DUNN, K. (1976). Kalman filter algorithms for a multi-sensor system. *Dans Decision and Control including the 15th Symposium on Adaptive Processes, 1976 IEEE Conference on*, volume 15, pages 570–574. IEEE.