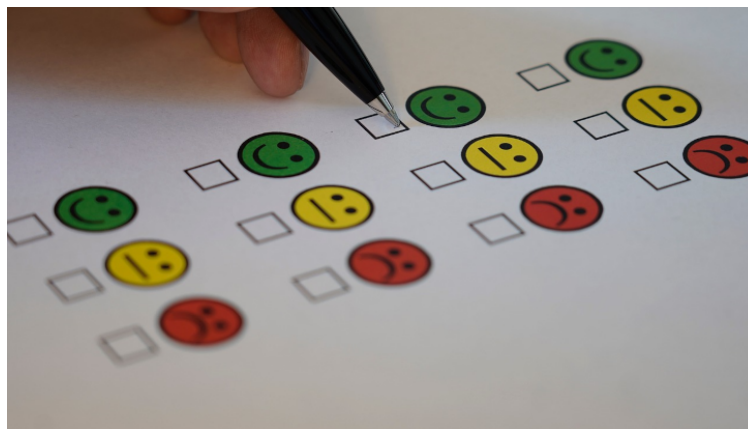


Divulga UAB

Revista de difusió de la recerca de la
Universitat Autònoma de Barcelona

23/09/2025

Quina etiqueta li assignem? Classificació en entorns d'incertesa



En l'àmbit de l'aprenentatge automàtic supervisat, la classificació és una tasca central que assigna etiquetes tenint en compte la seva distribució de probabilitats. Sovint es classifica segons el criteri de màxima probabilitat (MAP), però aquest pot ser inadequat quan les categories tenen un ordre, com en les escales de satisfacció. Aquest article proposa l'Ord-MAP, una alternativa òptima que obre la porta a una millor pràctica en classificació ordinal.

Classificar és una de les tasques centrals de l'aprenentatge automàtic supervisat, branca de la intel·ligència artificial que construeix, a partir de dades prèvies, sistemes capaços d'assignar etiquetes (classificadors) a nous ítems. Els classificadors no assignen etiquetes de manera determinista, sinó que ho fan en un context d'incertesa: per a cada nou ítem, proporcionen una distribució de probabilitats sobre les possibles etiquetes.

Quan les etiquetes són binàries (sí/no, bo/dolent...), és habitual assignar l'etiqueta amb la probabilitat més alta —la **moda** de la distribució. Aquest criteri, conegut com a **MAP** (Maximum A Posteriori), no és només intuïtiu, sinó també òptim si assumim que tots els errors de classificació tenen assignada la mateixa pèrdua, en el sentit que minimitza la pèrdua esperada quan classifiquem.

El criteri MAP també s'utilitza en classificació multiclasse nominal (més de dues etiquetes no ordenades), sota la mateixa hipòtesi de simetria en les pèrdues per errors de classificació. Tanmateix, en molts problemes reals, les etiquetes tenen un ordre intrínsec, com en les escales de satisfacció (molt insatisfet – insatisfet – neutre – satisfet – molt satisfet), conegudes com a escales de Likert. En aquesta situació, no tots els errors són igual de greus: classificar un client “molt insatisfet” com a “insatisfet” és menys greu que fer-ho com a “molt satisfet”.

Quan les etiquetes són ordenades, una alternativa més coherent al fet que tots els errors penalitzin igual seria tenir en compte la **distància** entre etiquetes. En un article recent, s'ha proposat un nou criteri de decisió per a la classificació ordinal: **Ord-MAP**, que consisteix a assignar la **mediana** de la distribució de probabilitats, és a dir, la primera etiqueta per a la qual la suma acumulada de probabilitats supera 0.5.

Per exemple, si un classificador retorna les probabilitats següents: 0.35, 0.05, 0.05, 0.30 i 0.25, corresponents a les cinc categories de satisfacció, el criteri MAP assignaria l'etiqueta “molt insatisfet” (la de major probabilitat), mentre que Ord-MAP escolliria “satisfet”, perquè és la primera categoria per a la qual la suma acumulada supera 0.5.

Aquest criteri, sorprenentment senzill, es demostra matemàticament a l'article que és **òptim**, ja que és el que minimitza la pèrdua esperada quan aquesta es defineix com la distància entre l'etiqueta real i l'etiqueta assignada. Els experiments realitzats amb diferents classificadors i bases de dades reals, i les simulacions, mostren clarament la superioritat del criteri Ord-MAP respecte del MAP fet servir habitualment.

Aquesta aportació obre la porta a una millor pràctica en classificació ordinal, tan present en aplicacions reals com les recomanacions automàtiques, les enquestes o l'avaluació de serveis.

Rosario Delgado

Departament de Matemàtiques
Universitat Autònoma de Barcelona
Rosario.Delgado@uab.cat

Referències

Delgado, R. (2025). **Ord-MAP criterion: Extending MAP for ordinal classification**. *Knowledge-Based Systems* 324, 113837, <https://doi.org/10.1016/j.knosys.2025.113837>

[View low-bandwidth version](#)