

Concept Drift

In the field of Machine Learning, Concept Drift is the actualized data being observed that displays statistical properties that vary from previously observed values or notions. This is an important topic to comprehend for systems that observe longer periods of data or that analyze volatile target data. Common fields of application where this problem persists include fraud detection, weather prediction, and online shopping customer behavior. There have been a slew of offered “solutions” to remedy the deteriorating predictive capacities of models ranging from:

1. Do nothing; more of a hands off approach and assume the data will never change. To me this seems like a naive approach, but nevertheless for quicker development times it may be required to make these types of assumptions
2. Re-fit the data; pretty simple solution, a good first pass at updating older models to newer data but still not a robust solution.
3. Vary data weights; the developer can implement a weighted system so that the impact of data is inversely related to its age
4. Learn the deltas; here we start to reach more complex solutions. Utilizing ensemble models you're able to spawn an additional model added to the ensemble of prior models to handle the newer changing data. Leading to a larger and larger ensembles as time progresses
5. Detect and adapt; here we dynamically change the deciding model based on the makeup of the data. So when the data does change so does the model to adapt to a volatile data space.
6. Data preparation; the ultimate pre processing of data over time. Employ varying pre-processing methods on data as time progresses so that the model can expect and predict on a baseline of data. This seems the most like neutralizing the change instead of adapting to it.

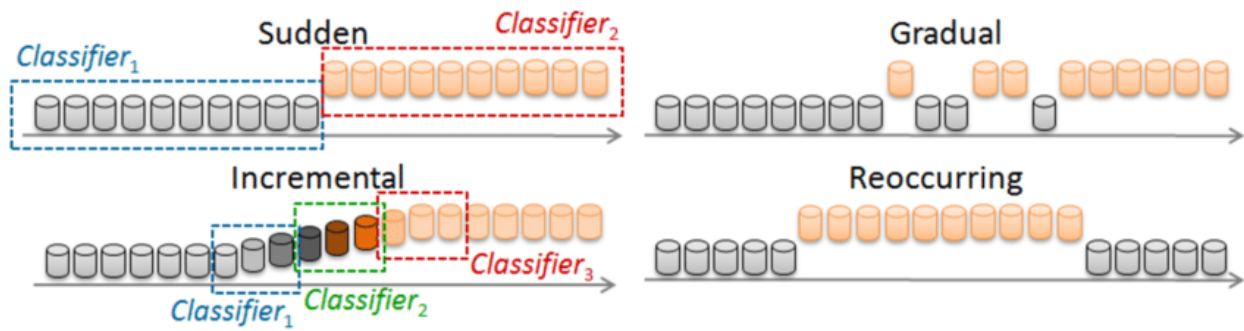
Some projects centered around the development of systems/models that can handle concept drift

- INFER: https://www.researchgate.net/publication/343879870_INFER_-_Computational_Intelligence_Platform_for_Evolving_and_Robust_Predictive_Systems
- GAENARI: <https://github.com/greenfish77/gaenari>
- ADEPT: <http://www.cs.man.ac.uk/~gbrown/adept/>

And the list goes on...

I will look into how these projects handle concept drift, what developments have changed over the past decade, and where the solutions are headed to overtake this problem.

Diagrams:



Sources:

<https://arxiv.org/abs/1010.4784>

<https://machinelearningmastery.com/gentle-introduction-concept-drift-machine-learning/>

https://www.win.tue.nl/~mpechen/publications/pubs/CD_applications15.pdf

<https://www-ai.cs.tu-dortmund.de/LEHRE/FACHPROJEKT/SS12/paper/concept-drift/tsymbal2004.pdf>