výsledného závěru některou z variant hlasování. Přehled různých metod lze nalézt v pracích (Dietterich, 2000) nebo (Bauer, Kohavi, 1999).

Při metodě zvané *bagging* (bootstrap aggregating) mají všechny modely rovný hlas. Předpokládejme použití jednoho algoritmu (např. rozhodovacích stromů). Pro potřeby učení se z daných dat vytvoří několik stejně velkých trénovacích množin pomocí náhodného výběru s opakováním (viz dříve zmíněná metoda bootstrap). Učení probíhá na jednotlivých trénovacích množinách nezávisle. Výsledné modely (rozhodovací stromy) se pak v procesu testování nechají hlasovat o přiřazení testovaného příkladu ke třídě.

Při metodě zvané *boosting* se postupně vytvářejí modely se stále větší váhou hlasu. Každý nový model je ovlivněn modely vytvořenými v předcházejících krocích. Ve fázi učení se nový model zaměřuje na ty příklady, které se nepodařilo zatím správně klasifikovat. Na počátku (při vytváření prvního modelu) mají všechny příklady stejnou váhu. Na základě výsledků klasifikace se váhy příkladů změní; u správně klasifikovaných se váha sníží, u chybně klasifikovaných se váha zvýší. To umožní od sebe odlišit „jednoduché" příklady (ty s nižší váhou) a „složité" příklady (ty s vyšší váhou). Celý postup se několikrát opakuje. Ve fázi klasifikace jednotlivé modely váženě hlasují o zařazení příkladu do třídy. Obrázek 139 ukazuje podobu tohoto algoritmu známou jako *AdaBoost* (Shapire, 1999). Podobně jako u předcházející metody i zde se předpokládá použití stejného algoritmu u všech modelů.

AdaBoost algoritmus

Učení

1. přiřaď stejnou váhu všem trénovacím příkladům,
2. pro každou iteraci (vytvářený model)
   2.1. vytvoř model
   2.2. spočítej chybu *Err* na vážených datech
   2.3. pokud $Err = 0$ nebo $Err \geq 0{,}5$ skonči
   2.4. pro každý příklad

   pokud klasifikace je správně, potom váha $w := w\dfrac{Err}{1-Err}$

   2.5. normalizuj váhy příkladů (součet nových vah stejný jako součet původních)

Klasifikace jednoho příkladu

1. přiřaď váhy 0 všem třídám
2. pro každý model

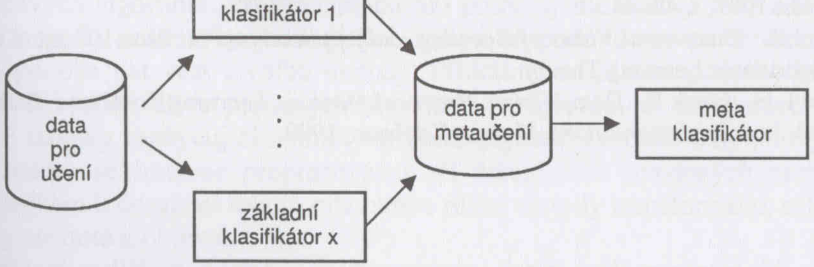
   přiřaď třídě určené modelem váhu $w := w - \log\dfrac{Err}{1-Err}$

3. vydej třídu s nejvyšší váhou

Obr. 139 Algoritmus AdaBoost.

V případě kombinování modelů vzniklých na základě různých algoritmů vyvstává jeden základní problém. Kterému modelu více věřit. Hlasování klasifikátorů má totiž smysl jen tehdy, jestliže všechny klasifikátory dávají srovnatelné výsledky. Pokud jsou ale některé modely výrazně horší, jejich zahrnutím do hlasování získáme horší výsledky než při použití jednoho lepšího modelu.

Otázkou jak rozpoznat spolehlivost jednotlivých modelů se zabývá metoda zvaná *stacking* (stacked generalization). Zavádí se zde koncept *metaučení,* což je použití metod učení ke zjištění jak nejlépe kombinovat jednotlivé modely. Tedy při metaučení se provádí učení z výsledků jednotlivých klasifikátorů. Vstupem do uvedené metody jsou výsledky klasifikace jednotlivými systémy. Počet vstupních atributů k metaučení (pro model na první úrovni) je tedy dán počtem modelů na nulté úrovni (základním klasifikátorem, base classifier); v nejjednodušším případě může být pro každý základní klasifikátor jeden atribut (indikátor odvozené třídy), složitější varianta může brát z každého základního klasifikátoru tolik atributů, kolik je tříd (např. pravděpodobnost zařazení objektu do těchto tříd). Cílovým atributem pro metaučení je skutečná třída. Výsledkem metaučení jsou tedy znalosti o tom, který základní klasifikátor se má použít (např. vždy věř modelu *A* a ignoruj model *B*). Klasifikace nových příkladů probíhá tak, že nejprve je příklad klasifikován základními klasifikátory a na základě těchto výsledků vydá model na první úrovni výsledné rozhodnutí. Schéma tohoto přístupu podle práce (Chan, Stolfo, 1993) je uvedena na obr. 140.



Obr. 140 Obecné schéma metaklasifikace.

Literatura

Bauer E., Kohavi R.: *An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants*. Machine Learning, 36(1/2): 1999, s. 105–139.

Berka P.: *Vybrané znalostní systémy, SAK, SAZE, KEX*. Skripta VŠE, Praha 1993, 245 s.

Berka P.: *Recognizing Reliability of Discovered Knowledge*. In: (Komorowski,-Zytkow, eds) Proc. Principles of Data Mining and Knowledge Discovery PKDD'97, LNAI 1263, Springer, 1997, s. 307–314.