Razpoznavanje objektov z metodo podpornih vektorjev *

Katarina Mele, Jasna Maver, Aleš Leonardis
Fakulteta za računalništvo in informatiko
Univerza v Ljubljani
Tržaška 25, 1001 Ljubljana, Slovenia
katarina.mele@fri.uni-lj.si, jasna.maver@fri.uni-lj.si, ales.leonardis@fri.uni-lj.si

Povzetek

V predstavljenem delu raziskujemo občutljivost metode SVM v primeru visokodimenzionalnih podatkov—slik. Metodo smo testirali na slikovni bazi COIL-100. Opazovali smo, kako se metoda obnaša pri slikah, ki smo jim dodali šum, zakrivanja in pri premiku predmetov na slikah. Rezultati so pokazali majhno občutljivost metode na šum in zakrivanja v slikah ter visoko stopnjo občutljivosti na premike predmetov na slikah.

1 Uvod

V sodobnih informacijskih sistemih se pojavlja vedno več zahtev po avtomatskem učenju in razpoznavanju predmetov v okolici. V zadnjem času narašča število aplikacij, ki za razpoznvanje vzorcev uporabljajo metodo podpornih vektorjev (angl. Support Vector Machine, okrajšano SVM). Ta se je doslej že uporabljala za razpoznavanje obrazov [4], oseb [2], ročno napisanih črk ter različnih vzorcev, ki jih srečamo v podatkih bioinformatike [1].

Osnovne zahteve sistema za razpozavo vključujejo zmožnost razpoznavanja predmetov iz različnih zornih kotov, pri različni osvetlitvi in pri nepopolni informaciji, ki je posledica zakrivanj. V delu analiziramo metodo SVM na bazi COIL-100, ki vključuje barvne slike zajete iz različnih smeri za sto predmetov. SVM metoda zahteva učno fazo. Pri učenju iz slik ni treba predhodno izločiti značilk — vhod v fazo učenja je lahko kar neobdelana množica slik.

SVM-metodo smo preiskusili s testno množico slik, ki se je od učne množice razlikovala. Razlike med množicama so posledica različnih smeri opazovanja, prisotnosti šuma in prisotnosti zakrivanj v slikah ter premikov predmetov na slikah v testni množici.

Matematična osnova za izračun SVM je predstavljena v drugem poglavju. V tretjem poglavju sta podrobno opisani faza učenja in faza razpoznavanja. Rezultati eksperimenta so predstavljeni v četrtem poglavju. Peto poglavje vključuje spoznanja in smernice za nadaljnje delo.

*Raziskava je potekala v okviru IST projekta EU CEC (IST-2000-29375)-CogVis, raziskovalnega programa *Computer Vision*-1539-506 ter projektov SLO-US-2002/13, SLO-A/07 in SLO-CZ 15/01-03, ki jih finančno podpira Ministrstvo za šolstvo, znanost in šport.

2 Metoda podpornih vektorjev

Metoda SVM je klasifikator, ki v osnovi razlikuje le med dvema razredoma. Mogoče pa jo je uporabiti tudi pri klasifikaciji več razredov. Pomembna značilnost metode je, da je uspešna pri klasifikaciji podatkov v visokodimenzionalnem prostoru značilk in kot taka primerna tudi za klasifikacijo slik. Podano množico, sestavljeno iz N slik velikosti $p \times r$, lahko prevedemo v množico vektorjev. Vsako sliko iz dane množice po stolpcih zložimo v vektor x_i velikosti n; $n = p \times r$. Označimo novo nastalo množico s simbolom S. Za točke množice S velja $x_i \in \Re^n$ i = 1, 2, ..., N. Vsaka izmed točk x_i sme pripadati le enemu izmed razredov $y_i \in \{-1, 1\}$. Metoda SVM poišče optimalno hiperravnino, ki razdeli množico S tako, da točke istega razreda ležijo na isti strani hiperravnine in pri tem maksimizira razdaljo med hiperravnino in njej najbližjimi točkami.

Pogoj za obstoj take hiperravnine je linearna ločljivost množice.

Definicija 1. *Množica S je linearno ločljiva, če obstaja $w \in \Re^n$ in $b \in \Re$, tako da velja*

$$y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \ldots, N. \tag{1}$$

Par (w, b) določa hiperravnino

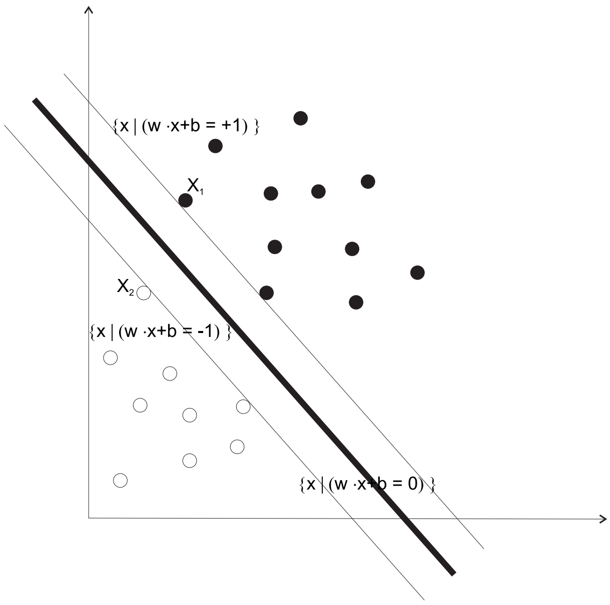
$$(w \cdot x_i + b) = 0. \tag{2}$$

Razdaljo točke x_i do hiperravnine (En. 2) določa naslednji izraz:

$$d_i = \frac{w \cdot x_i + b}{\|w\|}. \tag{3}$$

Definicija 2. *Optimalna hiperravnina je tista hiperravnina, ki je za dano linearno ločljivo množico S maksimalno oddaljena od najbližje točke iz S. Pri tem je razdalja najbližje točke do hiperravnine* *podana kot* $d = \frac{1}{\|w\|}$.

Definicijo 2 shematsko ponazarja slika 1.



Slika 1: Optimalna hiperravnina.

Iščemo torej tisto hiperravnino, pri kateri je d maksimalen. Ta hiperravnina ima najboljšo sposobnost generalizacije. Širino pasu, ki ločuje množici, lahko enostavno izpeljemo:

$$\begin{aligned}(w \cdot x_1) + b &= +1\\ (w \cdot x_2) + b &= -1\\ \Rightarrow (w \cdot (x_1 - x_2)) &= 2\\ \Rightarrow (\frac{w}{\|w\|} \cdot (x_1 - x_2)) &= \frac{2}{\|w\|}.\end{aligned} \tag{4}$$

Optimalno hiperravnino dobimo tako, da maksimiziramo razdaljo $2d$ (En. 4) pri omejitvi (En. 1), oziroma minimiziramo naslednjo enačbo:

$$\frac{1}{2} w \cdot w. \tag{5}$$

Naj označuje $\alpha = \alpha_1, \alpha_2, ...\alpha_N$ nenegativne Lagrangeove multiplikatorje izhajajoč iz omejitve (1). Rešitev zastavljenega problema je enakovredna določitvi sedla naslednje funkcije:

$$L_P(w, b, \alpha) = \frac{1}{2} ww - \sum_{i=1}^{N} \alpha_i (y_i(wx_i + b) - 1). \tag{6}$$

Rešitev zgornjega kvadratičnega problema poiščemo tako, da najprej poiščemo ustrezen dualen problem

$$L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j, \tag{7}$$

ki ga maksimiziramo. Dobljene vrednosti α uporabimo nato za izračun w-ja

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{8}$$

in nato s pomočjo Kühn–Tuckerjevih pogojev izračunamo še b

$$\alpha_i (y_i (x \cdot x_i + b) - 1) = 0. \tag{9}$$

α_i je različna od nič, če je pogojem iz enačbe (1) zadoščeno z enakostjo. Vektor w je torej linearna kombinacija tistih x_i, kjer je α_i različna od nič. Te točke so podporni vektorji in so izmed vseh točk iz S najbližje optimalni hiperravnini ter hkrati tudi zadostne za njen izračun.

Novo točko x_i klasificiramo tako, da pogledamo, kateri strani hiperravnine pripada:

$$sgn(w \cdot x_i + b). \tag{10}$$

Enačbe (5–9) so namenjene linearno ločljivim množicam S. Če množica S ni linearno ločljiva, enačbe posplošimo. Enačbo (1) popravimo tako, da pri iskanju optimalne hiperravnine dopuščamo tudi manjše napake:

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i, \qquad i = 1, 2, ...N \quad \xi \geq 0. \tag{11}$$

V našem primeru imamo opravka le z linearno ločljivimi množicami.

3 Opis poskusa

3.1 Testna baza

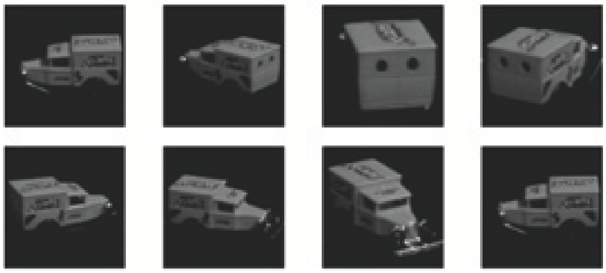
Metodo SVM smo testirali na bazi COIL-100 [3]. Bazo COIL-100 sestavljajo barvne slike stotih predmetov. Nekatere predmete prikazuje slika 2. Posamezen predmet je bil postavljen na rotacijsko mizico. Slike so zajete na vsakih 5 stopinj, kar da 72 slik za vsak predmet (slika 3).

3.2 Postopek učenja

Barvne slike RGB smo najprej pretvorili v sivinske slike s pomočjo naslednje enačbe

$$E = 0,31R + 0,59G + 0,10B. \tag{12}$$

Slika 2: Primeri objektov iz baze COIL-100.



Slika 3: Nekaj rotacij enega objekta.

Dobljene slike smo skalirali tako, da so sivinski nivoji zavzemali vrednosti od 0 do 255. Slike velikosti 128×128 smo pomanjšali na velikosti 32×32; pri tem smo povprečili kvadrate velikosti 4×4 slikovnih elementov. Tako pomanjšane slike smo ponovno sivinsko skalirali na območje med 0 in 255.

Slike enega predmeta so predstavljale en razred. Pri našem poskusu smo iz baze COIL-100 izbrali le 32 predmetov. Namenoma smo izbirali predmete, ki so si med seboj na pogled podobni (slika 2). Slike iz baze smo razdelili na učno in testno množico. 36 slik je pripadlo učni množici (z razmikom kota zajemanja 10 stopinj), ostalih 36 pa testni množici. Za 32 objektov smo tvorili vse možne pare in tako dobili 496 parov. Za vsak par predmetov smo z metodo SVM izračunali optimalno hiperravnimo. Tako smo dobili 496 optimalnih hiperravnin. Vsi pari učnih množic so bili linearno ločljivi.

Klasifikacijo v več razredov smo izvajali po principu teniškega turnirja. Vsak razred predstavlja enega igralca. V vsakem krogu sestavimo igralne pare. Z ustreznimi optimalnimi hiperravninami izračunamo zmagovalca vsakega para. V naslednji krog se uvrstijo le zmagovalci. Postopek ponavljamo, dokler ne dobimo absolutnega zmagovalca, to je razreda, v katerega objekt klasificiramo [5].

3.3 Faza testiranja

3.3.1 Slike brez popačenj

Najprej smo metodo SVM testirali na nešumnih testnih podatkih. Metoda je v tem primeru delovala zelo dobro, saj se je zmotila le pri eni izmed 1152 testnih slik. Tovornjak na sliki 4 je klasificirala kot gasilski avto prikazan v drugem stolpcu druge vrstice na sliki 2.

Slika 4: Napačno klasificiran objekt.

3.3.2 Slike z dodanimi popačenji

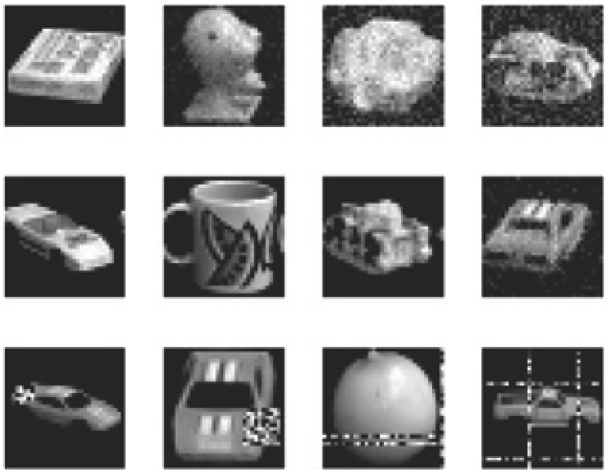
Metodo SVM smo testirali pri različnih popačenjih vhodnih podatkov. Slikam smo dodajali naključni šum, "premik" in zakrivanja.

Izbrali smo uniformni šum – vsakemu slikovnemu elementu dodamo naključno vrednost iz intervala $[-m, m]$. Uporabili smo uniformni šum z vrednostimi na intervalu $[0, 255]$. Premik k pomeni zamik stolpcev za k mest v levo. Zakrivanja na slikah so imela dve obliki – obliko kvadratov velikosti $l \times l$ in različnega števila vodoravnih in navpičnih črt. Pri tem so bile vrednosti elementov kvadratov in črt proizvedene naključno.

Eksperiment smo izvedli za naslednje kombinacije popačenj:

1. naključni šum,
2. premik,
3. naključni šum in premik,
4. zakrivanja v obliki kvadratov in
5. zakrivanja v obliki črt.

Prve tri vrste popačenj v eksperimentu smo dodali na sivinske slike velikosti 128×128. Slike smo nato sivinsko skalirali, pomanjšali in ponovno sivinsko skalirali. Zadnji dve vrsti popačenj pa smo dodali že pomanjšanim slikam. V teh dveh primerih smo slike najprej sivinsko skalirali, pomanjšali, jim dodali zakrivanja in nato ponovno sivinsko skalirali. Primeri slik s popačenji so na sliki 5.



Slika 5: Primeri šumnih slik.

4 Analiza rezultatov

Rezultate eksperimentov smo združili v tabeli 1.

Na nešumnih podatkih se metoda SVM obnaša zelo dobro. Klasifikacijska napaka je v primeru nešumnih podatkov le 0,09%. Z dodajenjem šuma in s širitvijo intervala se klasifikacijska napaka veča. Metoda se dobro obnaša pri majhnem šumu ($m \leq 50$), za ($m > 75$) pa postane nezanesljiva.

Metoda je zelo občutljiva na premike, saj se zanesljivost razpoznave zelo poslabša že pri manjših premikih ($k = 10$). Poudariti je treba, da so bili premiki narejeni na slikah velikosti 128×128.

V primeru zakrivanj s kvadratom se sistem obnaša robustno. Klasifikacijska napaka pri oknu velikosti 8×8 (na slikah 32×32) je le 6,32%. Prav tako so napake majhne pri prekrivaju s črtami.

	parametri	napaka (%)
brez šuma	—	0,09
	širina intervala	
šum	25	0,17
šum	50	5,12
šum	75	19,62
šum	100	32,73
šum	150	50,69
šum	200	64,93
šum	250	76,04
	velikost premika	
premik	3	1,39
premik	5	4,86
premik	7	12,93
premik	10	31,51
	šir. int., vel. pre.	
šum + pre.	25 3	0,95
šum + pre.	50 3	6,16
šum + pre.	100 3	32,55
šum + pre.	150 3	50,78
šum + pre.	25 5	5,12
šum + pre.	50 5	12,59
šum + pre.	100 5	38,54
šum + pre.	150 5	56,68
	velikost kvadrata	
kvadrat	4	0,43
kvadrat	6	1,91
kvadrat	8	6,34
kvadrat	10	16,75
	število hor. in vert. črt	
črte	1 1	2,08
črte	1 2	5,21
črte	2 1	3,39
črte	2 2	7,99

Tabela 1: Rezultati

5 Zaključek

Metoda SVM je učinkovita pri razpoznavanju predmetov na slikah. Sistem zelo dobro deluje na nešumnih podatkih, robusten je tudi pri manjšem šumu in zakrivanjih na sliki. Zelo je občutljiv že na manjše premike stolpcev na slikah, zato poseben izziv predstavlja nadgradnja metode SVM, ki bi omogočala robustno razpoznavo tudi v tem primeru.

Literatura

[1] N. Crisrianini and J. Shawe–Taylor:*An introduction to support vector machines and other kernel-based learning methods*. Cambridge University press 2000.

[2] C. Nakajima, M. Pontil and T. Poggio. People recognition and pose estimation in image sequences. *Proceedings of IEEE-INNS-ENNS International Joint Conference on Neural Networks*, Como, Italy, July 2000.

[3] S. A. Nene, S. K. Nayar and H. Murase: *Columbia Object Image Libary (COIL–100)*, Department of Computer Science, Columbia University, Technical Report No. CUCS–006–96.

[4] E. Osuna, R. Freund, F. Girosi: Training support vector machines: An application to face detection, *CVPR97*: 130–136, 1997.

[5] M. Pontil and A. Verri. Support vector machines for 3-D object recognition. *IEEE Trans. PAMI*, pp. 637–646, 1998.