

Obrazová detekce klíčových změn v sekvenci prezentačních snímků zaznamenaných systémem Mediasite

Bc. Martin Paroubek, Ing. Jan Silovský, Ph.D.

Abstrakt

Cílem této práce bylo vytvořit nástroj pro identifikaci duplicitních snímků v sadě podpůrných materiálů pro přednášky. Z dostupných materiálů byly vytvořeny dvě sady: trénovací a testovací. Extrahované vektory příznaků byly klasifikovány s použitím metody Support Vector Machines. Přestože byly nalezeny slabé stránky, byla metoda na základě výsledků navržena pro reálné použití, neboť v testovaných případech zachovala téměř všechny jedinečné snímky.

Úvod

Při nahrávání přednášek na Technické univerzitě v Liberci dochází k záznamu doprovodných textů přednášejícího (prezentace, psaní na tabuli, psaní rukou na papír). Výstupem ze systému je cca 1000 obrázků, ale pouze zhruba 3,5 % je jedinečných. Za jedinečné byly považovány nejobsáhlejší neduplicitní snímky.

Cílem této práce bylo vytvořit nástroj, spustitelný z příkazové řádky, který redukuje počet duplicitních snímků. Vstupem programu je adresář se sadou obrázků. Výstupem programu je seznam jedinečných snímků.

Ke splnění cíle byly ručně zpracovány dvě sady obrázků, každá o cca 10 000 snímcích, přičemž jedna slouží jako testovací a druhá jako trénovací. Všechny snímky byly rozostřeny a zmenšeny na jednotné rozlišení a změněny na černobílé. Poté byly použity metody k extrakci třiceti příznaků. Vzhledem k tomu, že určujeme míru duplicity, se konečnými příznaky staly součty absolutních hodnot diferencí příznaků dvou po sobě jdoucích obrázků. Mezi metody použité k extrakci patří histogram orientovaných gradientů, histogram stupňů šedi, hranová reprezentace a barva podle pixelu. Nakonec byla provedena klasifikace využívající Support Vector Machines (SVM).

Experiment a metody

Trénovací sada se skládala z 12 přednášek obsahujících 9 948 snímků a z toho 524 jedinečných. Testovací sada obsahovala 9 přednášek s 10 356 snímky a z toho bylo 201 jedinečných. Obě sady byly nasnímány ve dvou různých místnostech, a v obou sadách bylo kromě slajdů použito i focení tabule. První sada navíc obsahovala snímky zachycující kreslení přednášejícího na papír. Na těchto snímcích byl zachycen pohyb rukou a rotace papíru. Těmito negativními jevy se práce do hloubky nezabývala.

Hlavním klasifikátorem byly zvoleny Support Vector Machines (dále jen SVM). Wang [1] konstatoval, že v nejjednodušších případech SVM využívají nadrovinu lineárně oddělující data k vytvoření klasifikátoru s maximálním odstupem.

Úspěšnost byla měřena vypočítáním F-score (1). Při trénování s tímto ukazatelem je výsledkem nižší počet nesprávně identifikovaných jedinečných snímků.

$$F_{\beta} = (1 + \beta^2) \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}} \quad (1)$$

kde β je rovna 2 (ve prospěch jedinečných obrázků), precision je poměr správně klasifikovaných jedinečných snímků vůči všem klasifikovaným jedinečným snímkům a recall je poměr správně klasifikovaných jedinečných snímků vůči všem jedinečným snímkům.

Výsledky a diskuze

Pro všechny základní kernely SVM byly natrénovány hodnoty $cost$ a $gamma$, které posloužily k upřesnění chování. K tomu byl použit skript vlastní, který zkouší mocniny dvou pro zmíněné parametry $cost$ a $gamma$ (inspirováno nástrojem `grid.py` v `libsvm` [2]).

Tabulka 1. Výsledky pro natréované $cost$ a $gamma$ vlastním algoritmem

Kernel type	F_2 -score	Precision	Recall	TNR	$\log_2(C)$	$\log_2(gamma)$
Linear	90,21	100	88,06	100	-7	0
Polynomial	88,96	100	86,57	100	6	1
Radial basis f.	98,21	97,06	98,51	99,94	-5	6
Sigmoid	93,41	89,2	94,53	99,77	5	-5

Nejlépe dopadl SVM klasifikátor pro 30 příznaků s kernel RBF a parametry $\log_2 C = -5$ a $\log_2 g = 6$. Byl přitom vytvořen modul pro klasifikaci s 817 SV. Testovací sada byla tímto způsobem klasifikována s F_2 -score: 98,21 %. Na náhodně vybrané přednášce (mimo testovací sadu), která se ukázala jako jedna z nejsložitějších pro klasifikaci (z důvodu výskytu jevů, kterými se tato práce nezabývala), dosáhl F_2 -score: 67,16 % a recall: 100 %.

Výsledný program byl odzkoušen na sadě s 1841 snímky. Klasifikace snímků do tříd trvala 12 minut a 23 sekund. K měření byl použit notebook s 2 jádrovým procesorem Core 2 Duo 2,4 GHz a 2 GB RAM.

Závěr

Výsledný algoritmus dosáhl recall přes 98 %, ale v případě výskytu některých snímků (s rukou nebo stínem přednášejícího) docházelo k vysoké detekci FP a nižší hodnotě F_2 -score. Výsledný algoritmus byl problémový při určení příslušnosti slajdů, které se postupně zaplňují. Najít správnou hranici mezi těmito slajdy a například výstupem z konzole se ukázalo být problematické. Vzhledem k tomu, že program generoval nízké množství nesprávných detekcí jedinečných snímků je možné ho použít v praxi.

V navazující práci by mohlo dojít k vytvoření dalších algoritmů řešících: detekci snímků bez informační hodnoty (jednobarevné přechodové snímky, nebo rozmazané snímky), invarianci osvětlení a zakrytí. V případě sériového zapojení těchto algoritmů by mohlo dojít i k časové úspoře.

Poděkování

Rád bych poděkoval vedoucímu semestrálního projektu Ing. Janu Silovskému, Ph.D. za jeho rady a čas, který mi věnoval při řešení dané problematiky.

Reference

- [1] WANG, Lipo. *Support vector machines: theory and applications*. Berlin: Springer, 2005, s. 4. ISBN 3540243887.
- [2] HSU, Chih-Wei, Chih-Chung CHANG a Chih-Jen LIN. *LIBSVM: A library for support vector machines*. In: ACM Transactions on Intelligent Systems and Technology, 2011, volume 2, issue 3, s.1-27. Software dostupný z: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>