

Evaluace fuzzy ILP klasifikátoru na datech o dopravních nehodách

Jan Dědek, Peter Vojtáš

Katedra softwarového inženýrství, MFF, Univerzita Karlova v Praze,
Malostranské nám. 25, 118 00, Praha 1
`dedek@ksi.mff.cuni.cz`

Abstrakt Motivací pro tento příspěvek je klasifikace zpráv o dopravních a jiných nehodách podle jejich závažnosti. Pro tento účel jsme použili nový druh klasifikátoru, který je založen na induktivním logickém programování a fuzzy logice. V příspěvku nabízíme srovnání tohoto Fuzzy ILP klasifikátoru s dalšími alternativami, jmenovitě s rozhodovacími stromy J48, s klasifikátorem SVM a s vícevrstevným perceptronem. Výsledky tohoto srovnání ukazují, že Fuzzy ILP klasifikátor je na těchto datech srovnatelný až nepatrně úspěšnější než zmíněné alternativy.

Klíčová slova: fuzzy, ILP, klasifikace, strojové učení

1 Úvod

V našem výzkumu se zabýváme extrakcí informací z webových reportů. Jednou z datových sad, které používáme, jsou záznamy o hasičských výjezdech publikované na stránkách Hasičských záchranných sborů ČR¹. Klasifikace závažnosti nehody představuje přímočaré využití strukturovaných informací extrahovaných z reportu o ní. V tomto příspěvku se nezabýváme tím, jak získat strukturované informace z textového reportu (o tom viz např. naši práci [1]), ale tím, jak na základě takových informací určit závažnost dané nehody.

V naší předešlé publikaci [2] jsme představili nový druh klasifikátoru, který je založen na induktivním logickém programování a fuzzy logice.

Veškerý obsah této ukázky je formátován tak, jak je to požadováno pro Vaše příspěvky, tzn. slouží jako příklad, ze kterého můžete přímo vycházet. Doporučujeme jako výchozí použít tento soubor, který používá potřebné styly, které definují požadované rozměry tiskové plochy, dimenzování nadpisů, velikosti písma atd.

Veškeré formátování ani použití L^AT_EX-ovských konstrukcí není uváděno explicitně, zdrojový text této ukázky je samovysvětlující. Pro práci doporučujeme balík TexLive (Windows i Linux).

¹ <http://www.hzscr.cz/>

2 Pokyny pro autory

Příspěvek musí mít následující strukturu (v pořadí): název příspěvku, jména autorů², instituce a adresy autorů, email, abstrakt, klíčová slova, strukturovaný text, literatura, anotace v angličtině³.

2.1 Paragrafy, poznámky, matematické věty ...

Nadpis odstavce. Text pojmenovaného odstavce v rámci sekce. Odstavce nejsou číslovány a nadpis (ukončený tečkou) pokračuje přímo textem, není tedy na samostatném řádku.

Poznámka 1. Příklad poznámky.

Věta 1. *Text matematické věty. Podobně se uvádějí i definice, lemmata, důsledky apod.*

Důkaz. Důkaz věty ukončený čtverečkem. \square

Definice 1. *Text definice.*

Důsledek 1. *Text důsledku.*

Lemma 1. *Text lemma.*

2.2 Vzorce, příklady, tabulky, obrázky, ...

Vzorce, obrázky a tabulky musí být vycentrovány. Pokud je potřeba vzorec označit, používejte číslování na pravé straně, viz příklad.

$$L_p(v_x, v_y) = \left(\sum_{j=1}^n |v_x[j] - v_y[j]|^p \right)^{\frac{1}{p}} \quad (1)$$

Obrázky a tabulky formátujte podle následujících příkladů:

Tabulka 1. Popis tabulky (uvedený nad tabulkou).

Typ nadpisu	Příklad	Klíčové slovo
Název příspěvku	Ukázka použití ...	<code>\title</code>
Kapitola	Pokyny pro autory	<code>\section</code>
Sekce	Paragrafy, poznámky, ...	<code>\subsection</code>
Paragraf (odstavec)	Nadpis odstavce. Text pojmenovaného ...	<code>\subsubsection</code>
Poznámka	Text poznámky.	<code>\remark</code>

Obrázek může být uveden dvojím způsobem:

² Uvádění titulů u jmen podle Vašeho uvážení, kvůli přehlednosti jsou preferována jména bez titulů.

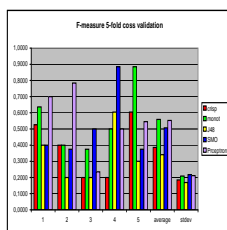
³ Pokud je příspěvek v češtině.

1. Obrázek, jehož obsah je buď text, nebo jehož grafika je zapsána přímo ve zdrojovém textu \LaTeX u, tj. obrázek, který není uložen v samostatném souboru EPS nebo PDF.

TEXT NEBO GRAFIKA OBRÁZKU

Obrázek 1. Popis obrázku (uvedený pod obrázkem).

2. Obrázek, jehož obsah je uložen v souboru EPS nebo PDF doporučujeme včlenit do textu pomocí makra `\image` ze souboru `unity.tex`, které rozpozná, zda je dokument překládán \LaTeX em, pak hledá soubor s příponou EPS, nebo je překládán $\text{PDF}\text{\LaTeX}$ em, potom hledá soubor s příponou PDF. Tímto způsobem lze udržovat jediný zdrojový text pro překlad oběma kompilátory. Samozřejmě je potřeba vygenerovat EPS a PDF soubor z obrázkem, EPS soubor lze jednoduše zkonvertovat do PDF pomocí utility `epstopdf.exe`, která je součástí balíku TexLive.



Obrázek 2. Popis obrázku z EPS nebo PDF souboru.

Příklad 1. Text příkladu.

2.3 Citace a reference

Všechny citace musí být uzavřeny v hranatých závorkách (např. [?]). Seznam referencí na konci příspěvku musí být uveden v abecedním pořadí dle autorů.

Formátování odkazu na článek v časopise nebo ve sborníku konference viz [?], formátování odkazu na knihu viz [?], formátování odkazu na www stránku viz [?]. Příklad viz [?].

Reference

1. Dědek, J., Vojtáš, P.: Linguistic extraction for semantic annotation. In Badica, C., Mangioni, G., Carchiolo, V., Burdescu, D., eds.: 2nd International Symposium

- on Intelligent Distributed Computing. Volume 162 of Studies in Computational Intelligence., Catania, Italy, Springer-Verlag (2008) 85–94
2. Dědek, J., Vojtáš, P.: Fuzzy classification of web reports with linguistic text mining. In Boldi, P., Vizzari, G., Pasi, G., Baeza-Yates, R., eds.: Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on. Volume 3., Los Alamitos, CA, USA, IEEE Computer Society (2009) 167–170

Annotation:

Title ...

Text ...