

prof. Ing. Jan Čapek, CSc.,
Ústav systémového inženýrství a informatiky,
Fakulta ekonomicko-správní, Univerzita Pardubice.

Autor: Ing. Marián Lamr
Školitel: doc. Ing. Jan Skrbek, Dr.
Studijní program: Systémové inženýrství a informatika
Studijní obor: Ekonomická informatika

Název práce: **Včasně varování před zvýšeným rizikem vzniku dopravní nehody s využitím data miningu.**
Posudek disertační práce.

Předložená disertační práce (DP) má 126 stran a je rozdělena do 7 kapitol a závěru práce. V první kapitole je popsán aktuální stav systémů zvyšující bezpečnost dopravy v ČR, tato kapitola je dále dělena na podkapitoly věnující se pasivním a aktivním prvkům plus telematickým systémům v ČR. Kapitola obsahuje jak systémy současně zaváděné do produkce automobilů (eCall) tak i systémy vzdálené budoucnosti např. Car2Car komunikace. Druhá kapitola je věnována data-miningu (DM) jeho technologii a DM nástrojům. Ve třetí kapitole jsou popsány vybrané algoritmy DM. Čtvrtá kapitola popisuje zdroje dat o dopravních nehodách a alternativní možnosti jejich využití. V páté kapitole je představen cíl DP. Šestá kapitola popisuje principy systému včasného varování před dopravní nehodou, sedmá kapitola se soustřeďuje na hledání skrytých závislostí v datech o dopravních nehodách pomocí DM nástrojů jako součást řídicího systému včasného varování. Závěr obsahuje shrnutí DP. Dále je uveden seznam literárních zdrojů a publikace autora související s tématem DP.

Dále se vyjádřím k:

- Aktuálnosti zvoleného tématu disertační práce.
- Splnění sledovaného cíle.
- Zvoleným metodám zpracování disertační práce.
- Dosaženým výsledkům disertační práce
- Přínosům pro další rozvoj vědy a techniky

Aktuálnost zvoleného tématu disertační práce.

Není pochyb o tom, že zaměření na zvýšení bezpečnosti silniční dopravy je velice aktuální. S pokroky v informačních a komunikačních technologiích a s rostoucí hustotou provozu na pozemních, ne příliš kvalitně udržovaných, komunikacích nabývají systémy včasného varování na důležitosti.

Splnění sledovaného cíle.

Po prostudování DP mohu prohlásit, že dizertantem vytčený hlavní cíl- vytvořit konceptuální návrh systému umožňujícího v reálném čase a místě predikovat riziko dopravní nehody – a dále odvozené podcíle, které jsou podrobně popsány v páté kapitole, jsou splněny.

Zvolené metody zpracování dizertační práce.

Dizertant při zpracování DP použil základní vědecké metody, nejprve analýzou současného stavu řešené problematiky zjistil slabá místa, která se především týkala zjištění, že se pro on-line predikci rizik vzniku dopravní nehody nevyužívají databáze dopravních nehod. Z toho potom vyplynula řešení problému.,

Dosažené výsledky dizertační práce s ohledem na nové poznatky.

Za přínos práce lze považovat zejména ideové řešení problematiky včasného dopravního varování s využitím databáze policie ČR o dopravních nehodách.

Přínosy pro další rozvoj vědy a techniky.

Využitelnost výsledků DP je ve smyslu tohoto odstavce v ideovém řešení problematiky včasného dopravního varování s využitím databáze policie ČR o dopravních nehodách, vytvářením predikcí rizik vzniku dopravní nehody.

Připomínky k dizertační práci:

Druhá a třetí kapitola by mohly být stručnější, měly by se věnovat jen technikám použitým v práci.

Str. 77 Věta: *Aplikace v automobilu porovnává výsledek predikce s aktuální situací a ve chvíli kdy podobnost překročí předem stanovenou hladinu je řidič upozorněn.* Toto předpokládá (a v práci je to několikrát zmíněno) že aplikace v automobilu je spojena se všemi relevantními snímači, kterými je automobil vybaven. Za současného stavu jde jen o ideový návrh praxi hodně vzdálený.

Str. 77 není popsán systém RADIA HELP.

Str. 97 V tabulce 4 chybí atribut zohledňující kvalitu komunikace, Sice na str. 94 autor zavádí atribut *Stav vozovky v době nehody* ale to myslí jen z pohledu počasí. Nikde se neuvažuje stav vozovky z pohledu kvality vozovky.

Str. 102 V kapitole se operuje s pojmem hustota dopravy a bere se průměrná hodnota za 24 hodin. Přitom je známo, že se hustota dopravy v různých denních i nočních hodinách

od průměrné hodnoty výrazně liší. Hustota dopravy má významný vliv na vznik rizika dopravní nehody.

V práci mě chybí porovnání jednotlivých metod DM zmíněných v DP, co do času zpracování a přesnosti stanovení shluků.

V práci není brán zřetel na současný stav vozového parku v ČR.

Za jakých podmínek by Váš ideový návrh byl prakticky použitelný?

Závěr:

Není pochyb o tom, že zaměření na zvýšení bezpečnosti silniční dopravy je velice aktuální, zvláště v současné době, kdy počty automobilů na pozemních komunikacích rostou geometrickou řadou. I přes mé výhrady doporučuji předloženou DP k obhajobě a po úspěšné obhajobě

n a v r h u j i

Ing. Mariánu Lamrovi udělit akademický titul Philosophiae Doctor ve zkratce Ph.D.

V Pardubicích dne 15. 10. 2018

Oponentní posudek disertační práce

Název práce: VČASNÉ VAROVÁNÍ PŘED ZVÝŠENÝM RIZIKEM VZNIKU DOPRAVNÍ NEHODY S VYUŽITÍM DATA MININGU

Autor práce: Ing. Bc. Marián Lamr

Školitel: doc. Ing. Jan Skrbek, Dr.

Oponent: prof. RNDr. PhDr. Antonín Slabý, CSc., Fakulta informatiky a managementu, Univerzita Hradec Králové

Struktura a obsah práce

Práce má 126 stran včetně seznamů a příloh. Základní text je kromě úvodu a závěru členěn na 7 číslovaných kapitol.

Nečíslovaný Úvod stručně představuje problematiku dopravních nehod v základních statistikách v posledních 10-12 letech a z toho vyplývající důležitost tematiky a nových přístupů k řešení problematiky a naznačuje strukturu a cíle práce.

Kapitola 1 - Aktuální stav systémů zvyšujících bezpečnost dopravy v ČR - shrnuje současně využívané pasivní prvky a aktivní prvky bezpečnosti a telematické systémy v ČR. V rámci telematických systémů je představen zejména RDS-TMS a eCall.

Kapitola 2 - Data mining: metodologie, big data, základní přístupy a typické úlohy - je zaměřena na vybrané technologie data miningu, jež jsou adekvátní pro aplikační oblast a vybraným algoritmům se zdůrazněním shlukové analýzy, a dále i softwarovým produktům ke zpracování dat z aplikační oblasti.

Kapitola 3 - Vybrané algoritmy data miningu - je věnována algoritmům, připadajícím v úvahu pro řešenou problematiku. Jde např. o asociační pravidla a algoritmus Apriori a zejména algoritmy z oblasti shlukování (K-means, DBSCAN, OPTICS, DENCLUE a v práci použitý TwoStep (implem. IBM SPSS Modeller).

Kapitola 4 - Zdroje dat o dopravních nehodách a alternativní možnosti jejich využití - se zabývá databázemi o dopravních nehodách a klasifikací dat a potenciálu využití těchto zdrojů dat.

Kapitola 5 - Cíle práce, metody a další směřování disertační práce - formuluje cíle, kterých chce autor dosáhnout, nastiňuje metody a formuluje přínosy práce.

Kapitola 6 - Princip systému včasného varování před dopravní nehodou - se detailně zabývá autorovým nejpodstatnějším zamýšleným cílem, výsledkem práce, prvkům a vazbám tohoto systému a vybraným modelovým situacím vedoucím k jeho použití.

Kapitola 7- Hledání skrytých závislostí v datech o dopravních nehodách pomocí DM nástrojů jako součást řídicí části systému včasného varování - ukazuje na dvou úlohách detailně způsoby analýzy dat a problému včetně navazujícího modelování.

Nečíslovaný Závěr rekapituluje obsah doktorské disertační práce, dosažené snažení a výsledky a formuluje i další možné rozvinutí tématu.

Následují obligátní: Seznam použité prameny a Publikace autora, související s tématem disertační práce.

Cíl práce, aktuálnost tématu

Hlavním cílem práce je vytvoření konceptuálního návrhu systému, umožňujícího v reálném čase a místě predikovat riziko dopravní nehody. Systém je založen na využívání predikčních modelů, které jsou vytvářeny pomocí data miningových technik a nástrojů, využívajících především algoritmy shlukové analýzy a asociačních pravidel.

Díličmi cíli práce jsou:

- Analýza současného stavu systémů zvyšujících bezpečnost účastníků silničního provozu
- Analýza využitelnosti technik dobývání znalostí z databází
- Určení vhodných algoritmů a postupů v souvislosti s realizací daného systému
- Návrh a implementace vlastní databáze dopravních nehod.

Cíli práce je i dosažení výsledků metodických:

- Systémový pohled na problematiku,
- Vytvoření modelů schopných identifikovat nebezpečná místa na silničních komunikacích.

Hlavní cíl je inženýrský, náročný. Model vyžaduje využití/zpracování různě strukturovaných a různě úplných dat. Přístupy a metody v rámci tohoto cíle jsou z hlediska struktury klasické, z hlediska nutnosti reakce na speciální problematiku a aplikační oblast dosti specifické a náročné, mající slušný výzkumný a aplikační potenciál. Jde o tematiku a cíle, jež jsou moderní, nové, náročné, zajímavé, v literatuře nezpracované. Původnost tématu a cílů a výběr náročné aplikační oblasti je třeba velmi pozitivně hodnotit. Téma je v souladu se zaměřením doktorského studia oboru Ekonomická informatika doktorského studijního programu Systémové inženýrství a informatika.

Použité vědecké metody v disertační práci

Práce vychází z rešerše a analýzy obtížněji dostupných, dosti roztržitých a neúplných literárních zdrojů o předmětné problematice (zejména jde o způsoby aplikace obecných postupů na zkoumanou oblast). Zásadní pozornost je věnována způsobům získání relevantních dat a informací z veřejně dostupných zdrojů.

Autor použil metody a postupy, které je možno považovat za adekvátní, osvědčené a ke splnění konkrétních cílů vedoucí. Kromě obecných metod a postupů vědecké práce, kde dominuje snaha systémový přístup, patří mezi hlavní metodické postupy ke zpracování dat zamýšlené v práci metody patřící do oblasti dataminingu - klasifikace a predikce, analýza vztahů, seskupování, analýza časových řad, detekce anomálií. Dominují metody a algoritmy shlukování - (metody rozdělovací, hierarchické, založené na hustotě, mřížkové). V práci jsou z hlediska použitelnosti podrobně zkoumány vybrané algoritmy Algoritmy (K-means, DBSCAN, OPTICS, DENCLUE a v práci použitý TwoStep (implementace IBM SPSS Modeller). Autor prozkoumal funkčnosti a

možnosti vybraných softwarových produktů, zejména Soft nástroje SPSS Modeler, KNIME, RapidMiner studio, Orange, Weka, ELKI. Zamýšlené metodické postupy práce se rovněž jeví jako adekvátní a tedy rozumné.

Splnění cílů práce

Cíle práce byly splněny. K hlavnímu cíli práce a podobně ke všem dílčím krokům vedoucím k jeho dosažení autor dostatečně přispěl. Metodický rámec se jeví jako použitelný, využitelný a dále rozpracovatelný.

Přesnost práce, formální stránka práce

Práce je napsána jasným a přesným jazykem. Formální stránka práce i přesnost vyjadřování, úprava vzorců, diagramů obrázků a výstupů jsou na dobré úrovni. Práce má jasnou strukturu, proto je možno se v ní dobře orientovat. Jazykové prohřešky a překlepy se vyskytují velmi omezeně.

Výsledky práce a poznatky a přínosy práce

Aplikační oblast je dosti specifická, značně náročná a na získání dat ale i přesnost postupů citlivá. Výsledky mohou být proto využity zejména v oblasti testování dat, rozboru výsledků, návrhů a doporučení.

Nelze očekávat hotová řešení anebo dokonce již do praxe nasazená řešení, jako je tomu v jiných oblastech aplikací

Práce rovněž přináší výsledky v oblasti teoreticko-metodologické. Za hlavní přínos je možno považovat vytvoření a popis konceptuálního návrhu systému, umožňujícího v reálném čase a místě predikovat zvýšené riziko dopravní nehody. Popis jeho funkčností a komponent, využívajících predikčních modelů s bohatým využitím data miningových technik a nástrojů využívajících především algoritmy shlukové analýzy a asociačních pravidel, dává řadu podnětů v oblasti funkčností těchto systémů.

Práce může mít i rozsáhlé využití praktické. Nabízí se opakování postupů uvedených v práci, další rozšíření a zobecnění metodiky a praktická doporučení týkající se např. definice struktury sbíraných dat s cílem efektivnějších řešení včasného varování, predikce rizika vzniku nehody apod., využití systému eCall a pod., a využití výsledků na příbuzné aplikační oblasti (kriminalita apod). Samostatnou hodnotu mají i 2 komplexní případové studie.

Publikace autora ve vztahu k práci

Publikační činnost autora má 19 položek a z toho je 11 věnováno tématům blízkým tématu doktorské disertační práce. Publikační činnost ukazuje autorovu tematickou vyhraněnost, publikační spolupráci se školitelem, návaznost na výzkumné aktivity katedry. Autor v disertační práci poměrně unikátně využívá všech svých dosavadních zkušeností.

Otázky do diskuse obhajobě

Práce ukázala zároveň, že problematika je náročná a ne úplně přehledná s velkými obtížemi v oblasti přenosu dat a interoperability mezi sítěmi a dosažení standardů.

Diskuse by se proto měla týkat hlavního autorova výsledku, tj. podstaty řešení, problémům, které musel autor překonat, rozsahu testování a zkušeností s jejím použitím, jejím možným dalším vývojem/rozšířením.

Lze navrhovanou metodiku využít i pro jiné oblasti prevence?

Rozprava by se mohla věnovat i upřesnění některých pojmů obecných či speciálních, jako např.:

Jak by bylo možno dále postupovat od systémového pohledu k tvorbě komplexního software a jeho dokumentaci?

Můžete popsat důvody a případná úskalí zavedení parametru hustoty dopravy do datové matice a v návaznosti do celého navrhovaného řešení?

Závěr:

Práce splňuje nároky na disertační práce kladené. Autor není v oblasti tématu práce nováčkem a přispěl k jejímu rozvoji svojí disertační prací i dalšími publikacemi věrohodným příspěvkem. Navázal přitom i na výsledky práce pracoviště a školitele. Doporučuji, aby Ing. Mariánu Lamrovi byl po úspěšné obhajobě udělen titul Ph.D.

Hradec Králové, 14. 10. 2018

Antonín Slabý