

CLUSTERING OF ECG CYCLES

Richard Ředina

Bachelor Programme (3rd year), FEEC BUT

E-mail: xredin00@stud.feec.vutbr.cz

Supervised by: Marina Ronzhina

E-mail: ronzhina@feec.vutbr.cz

Abstract: The study is focused on a design of a reliable approach for ECG cycles clustering. It would be helpful for automatic assessment of various pathological patterns in ECG. Proposed method was tested and tuned on real data from ambulatory ECG database. The algorithm comprises ECG preprocessing, adjustment of R-peak positions available in database, creation of a template cycle, computation of features mainly representing correlation between particular cycles and the template, and, clustering of cycles within ECG via k-means. The appropriate number of clusters is derived via analysis of silhouette values. Resulting success of the algorithm in comparison with available manual scoring is: Sensitivity = 0.55 and Specificity=0.94.

Keywords: ECG, QRS complex, cardiac arrhythmias, cluster analysis, k-means, correlation, silhouette

1 ÚVOD

I přes fakt, že je srdce životně důležitý orgán, podléhá postupem času různým patologiím, které jsou následně spojeny s pozměněnou funkcí. Velmi často jsou různá poškození srdce odražena v jeho elektrické aktivitě, která je snímána jako EKG [1]. Včasné rozpoznání patologií v pacientově záznamu může zlepšit jeho prognózu. Při automatickém zpracování je tedy potřeba vytvořit algoritmus, který bude záznam vyhodnocovat a poskytovat údaje o srdeční aktivitě, jež je neobvyklá. Nástrojem pro vyhodnocování může být právě shluková analýza.

2 DATABÁZE EKG

Ústavem biomedicínského inženýrství byl pro tuto práci poskytnut přístup k databázi EKG signálů shromážděné firmou BTL[®]. Databáze obsahuje celkem 6884 záznamů, které byly předzpracované automatickým softwarem. Součástí databáze je tedy i samotné rozměření signálu, které ovšem v některých případech není dokonalé. Do jednotlivých záznamů se zároveň promítalo rušení, a to jak z oblasti nízkých frekvencí, tak i z oblasti 50 Hz. Tyto skutečnosti jsou řešeny v rámci předzpracování. Záznamy jsou složeny celkem z osmi svodů (II, III, V1-6) a také krátké anotace k rytmu a případným patologiím.

3 SHLUKOVÁ ANALÝZA

Jak plyne z názvu, nástrojem pro rozdělování shluků je shluková analýza [2], tedy postup hledání skupin objektů, jež jsou si svými vlastnostmi podobné a zároveň jejich jasné oddělení od skupin, se kterými mají jen málo společného. Konkrétně je použita metoda *k-means* [3], kterou lze považovat za pilíř celé práce. Vstupem je sada příznaků reprezentujících jednotlivé srdeční cykly (úseky kolem R vln, viz dále). Příznaky jsou odvozeny na základě srovnání se vzorem (umělým standardem), jež bude vytvořen kumulací hodnocených cyklů v záznamu. Cykly jako takové jsou vybírány ze záznamu jako úseky kolem detekovaných R kmitů. Velikost cyklů je závislá na srdeční frekvenci. Je vybrán nejkratší RR interval ze záznamu, jeho délka je vydělena dvěma a od výsledku je odečtena konstanta (volitelná, ale ideálně co nejmenší). Získané číslo charakterizuje okolí kmitu R na obě

strany. Cykly se nepřekrývají. Soubor příznaků bude sestávat ze střední kvadratické odchylky, korelace (obojí jak v časové oblasti, tak v oblasti spektrální) a také hodnoty maximální výchylky ve srovnávaném úseku.

Kvalita proběhlého shlukování je hodnocena tzv. *siluetou*, která charakterizuje míru správného přiřazení objektu do shluku. [4]

4 ZPRACOVÁNÍ

Práce se záznamy pacientů se skládala z předzpracování, vytvoření vzoru, shlukování a následného hodnocení. Jednotlivé kroky nyní rozeberme podrobněji.

4.1 PŘEDZPRACOVÁNÍ

Výše bylo uvedeno, že obdržená databáze nebyla v některých ohledech zcela ideální. Prvním krokem předzpracování byla filtrace filtrem typu HP a DP s mezními frekvencemi 1 Hz a 40 Hz. To ze signálu odstraní kolísání nulové linie a eliminuje síťový brum. [5]

Dalším krokem je úprava pozic R kmitů, které nebyly detekovány zcela přesně a pro následné porovnávání je důležité, aby byly signály zarovnané. Pozice detekované firmou BTL nejsou však příliš daleko, od skutečných maximálních výchylek, proto poslouží jako dobrý odrazový můstek. Úprava spočívá ve vybrání lokálního maxima (eventuálně minima) v každém cyklu.

V EKG se však místy objeví situace, kdy kladný i záporný kmit jsou přibližně stejně velké. V tom případě by se občas vyhodnotilo R kladné a místy záporné, což by nespĺňovalo požadavek zarovnání. Kvůli tomu byl do algoritmu přiřazen krok, který vyhodnotí, zdali je vyšší počet kladných nebo záporných výchylek a následně ověří, jestli se u minoritní skupiny neobjevuje i opačná výchylka. Pokud ano a její absolutní hodnota se příliš neliší od hodnot majoritních výchylek, bude vzata jako R kmit.

Z takto upravených R kmitů jsou odvozeny délky jednotlivých cyklů. Následně se vytvoří kumulací vzor a vypočtou se příznaky charakterizující jednotlivé srdeční cykly.

4.2 SHLUKOVÁNÍ SRDEČNÍCH CYKLŮ

Jak bylo řečeno výše, ke shlukování je použita metoda *k-means*, která si žádá znát předem počet shluků, který však z patientských snímaní není k dispozici. Zde je zaveden předpoklad, že v rámci jednoho deseti vteřinového záznamu se nevyskytne více jak $N/2$ různých cyklů, kde N je celkový počet cyklů do analýzy vstupující. Z toho tedy vyvodíme, že počet shluků, pro které budeme provádět metodu *k-means* bude 2 až $N/2$. Po každém proběhlém shlukování je vypočtena průměrná hodnota siluety. Pakliže dosáhne hodnota siluety 0.9 a vyšší nebo 0.8 s podmínkou, že žádná dílčí hodnota siluety neklesla pod 0.75, je velice pravděpodobné, že shlukování je pro odpovídající n shluků vhodné. S nižší hodnotou tato pravděpodobnost klesá. Není však možné ji zcela vyloučit. Z toho důvodu je potřeba provést statistické zhodnocení metody a následně poupravit práh pro vyhodnocení správného shlukování.

5 STATISTICKÉ ZHODNOCENÍ

Pro zhodnocení bylo vybráno náhodně 500 záznamů z databáze BTL[®]. Takové množství je dostatečně velké na to, aby mělo vypovídající hodnotu a zároveň se do nich promítly jak záznamy fyziologické, tak s patologiemi. Do analýzy nebyly zahrnuty záznamy, které obsahovaly artefakty způsobené kardiostimulátory. Záznamy byly zpracovány a výsledky byly porovnány s přiloženou anotací. V tabulce 1 jsou zapsány četnosti výsledků. V řádcích tabulky je zachyceno, zdali byl (+) nebo nebyl (-) v anotaci zapsán abnormální komorový komplex nebo zvláštní morfologická odchylka v oblasti P vlny. Sloupce odpovídají výsledkům vytvořeného algoritmu, tedy jestli rozřadil srdeční cykly do více shluků (+) či nikoli (-).

Tabulka 1: Kontingenční tabulka

		Algoritmus		Σ
		+	-	
Anotace	Σ	147	353	500
	-	65	333	398
	+	82	20	102

Z tabulky 1 je patrné, že celkem 85 záznamů bylo nesprávně vyhodnoceno. Případy, kdy se abnormality vyskytovaly, ale nebyly detekovány (chyba I. řádu), byly pravděpodobně způsobeny nedostatečnou citlivostí na odlišení pozměněných cyklů. Pro snížení této chyby by bylo vhodné přidat do hodnocení příznaky, které nesou informaci odlišnou od ostatních a tím je i vyhodnotit jako rozdílné. Chyba II. řádu, kdy algoritmus vyhodnotil cyklus jako odlišný byla často způsobena nesprávným zarovnáním QRS komplexů, které vzniklo pravděpodobně během úpravy pozic R vln. Eliminovat by šla přesnějším vyhodnocením kmitu R.

Z kontingenční tabulky mohou být dopočítány další parametry, a to konkrétně specifita (TNR), senzitivita (TPR), pozitivní prediktivní hodnota (PPV) a negativní prediktivní hodnota (NPV).

Tabulka 2: Dopočítané parametry

TNR	0.94
TPR	0.55
PPV	0.8
NPV	0.83

Zanesené hodnoty v tabulce 2 ukazují, že algoritmus je poměrně specifický, bohužel už ne tolik senzitivní. Prediktivní hodnoty, které poskytují informaci o pravděpodobnosti správného zařazení záznamu jsou poměrně dobré.

Algoritmus jako takový je citlivý na správnost detekce R kmitů. Vzhledem k tomu, že v rámci zpracování se berou v úvahu všechny dostupné pozice R kmitů z databáze, tak ani dodatečná úprava pozic nebude efektivní, pokud je detekována například vlna T místo QRS komplexu. Takovýchto případů bylo v rámci těchto pětiset záznamů celkem 13. Pokud bychom tedy předpokládali dokonalou detekci R kmitů, a tím bychom eliminovali těchto 13 záznamů, zvedla by se senzitivita na 0.61 a NPV na 0.86.

6 ZÁVĚR

Celý algoritmus lze označit jako poměrně specifický, zároveň má celkem dobré PPV a NPV, které označují pravděpodobnost, správného rozřazení do shluků. Senzitivita jako taková není zcela ideální. Nespornou výhodou algoritmu je, že mu stačí detekované pozice R kmitů, což je v dnešní době poměrně rutinní záležitost. Správnost poloh R vln je však klíčová a výsledek shlukování na ní závisí.

Algoritmus jako celek by mohl v praxi sloužit pro automatické sledování a hodnocení delších srdečních záznamů (např. z denního holterovského měření). Výstup jasně shrnuje podobu a četnosti jednotlivých srdečních cyklů.

PODĚKOVÁNÍ

I would like to thank Ing. Marina Ronzhina Ph.D. for her advices. I am also very grateful for her helpfulness. At the same time, I want to thank for providing an access to a test database. The entire algorithm would be difficult to create without it. Eventually I thank my schoolmates for their support.

REFERENCE

- [1] CURTIS, Michael J., Jules C. HANCOX, Andres FARKAS, et al. The Lambeth Conventions (II): Guidelines for the study of animal and human ventricular and supraventricular arrhythmias. *Pharmacology and Therapeutics*, Amsterdam: Elsevier Inc., 2013. roč. 139, vyd. 2, s. 213-248. ISSN 0163-7258
- [2] RONZHINA, Marina. Shluková analýza. Presentation presented at: [Umělá inteligence v medicíně AUIN; 2018 Oct 23, Brno, Czech Republic]
- [3] DUDA, Richard O., Peter E. HART, David G. STORK. *Pattern Classification*. New York, United States: John Wiley and Sons Inc., 2000 ISBN: 0-471-05669-3
- [4] RONZHINA, Marina. Nehierarchické metody. Presentation presented at: [Umělá inteligence v medicíně AUIN; 2018 Oct 30, Brno, Czech Republic]
- [5] KOZUMPLÍK, Jiří. Analýza biologických signálů [online prezentace]. Brno: ÚBMI, VUT, [cit. 2. 12. 2013]