

「異種センサーデータによる個人健康状態の推定と応用について」

戴 イン (ソフトウェア情報学部、准教授)

<要旨>

今年度の研究では、主に、主観的領域における分類器の性能を改善するために、K平均クラスタリングを用いて主観的にラベル付けされたトレーニングデータの品質を測定するための基準(QoSTD)を定義した。次に、クラスの予測スコアの重みとして QoSTD が使用され、推測されたインスタンスの可能性を調整する。さらに、主観的なラベル付きデータによって訓練された分類器の性能を評価する。提案された方法の有効性を検証するために、support vector machine (SVM) と discrimination analysis (DA) モデルを利用して、提案された方法によって、東洋医学に関わる 13 個の基本証を推測する分類器が実世界の学習データから訓練された。実験結果は、インスタンスに対する観測値とのその尤度の一致率が、特に比較的低い QoSTD を持つ訓練データセットの場合に、クラスごとで顕著に増加することを示した。実験結果はまた、レーニング・データ・セットから間違いラベル付けされてインスタンスを除去し、サブジェクト・ドメイン内の分類器を再トレーニングする方法と解決策を示した。それにより、提案された方法が個人健康状態を推定する分類器の性能を改善することを検証した。

1 研究の概要

バイオセンサーを含むセンサー技術の進歩は、健康管理分野への新たな応用を可能にした。血圧計や心電計などボディセンサーを使って、日常生活の中の人の健康状況を把握する健康管理システムが注目されている。しかし、現在のシステムは健康状態の異常をチェックできるが、異常の原因が確認できないため、健康状態の改善には至らない。また、対象者への負担を考慮すると、体につけるセンサーを少なくする必要があるが、適切なセンサーを選定する基準が明確になっていない。

一方、東洋医学では東洋医師が対象者の健康状態を四診(望診、問診、聞診、切診)により、虚/実と熱/寒と裏/表の3次元の原点から偏位した基本13証(寒、熱、気虚、気滞、血虚、血瘀、津虚、痰飲、心、肺、脾、肝、腎)に基づいて判断し、患者それぞれの体質に合わせて、症状を改善する養生(食事などの工夫)と治療法を提供する。東洋医学の病理概念には、病気というほどではないが放置すると病気に至るという未病といわれる状態があり、この段階で養生を取り入れることで、発病を予防できるとされている。

本研究では、東洋医学の基本証を用いて効率的で精度の高い健康状態推定モデルの構築を目指し、利用者の健康管理と増進への活用する仕組みを検討する。

2 研究の内容

本研究の流れが図1の通りである。

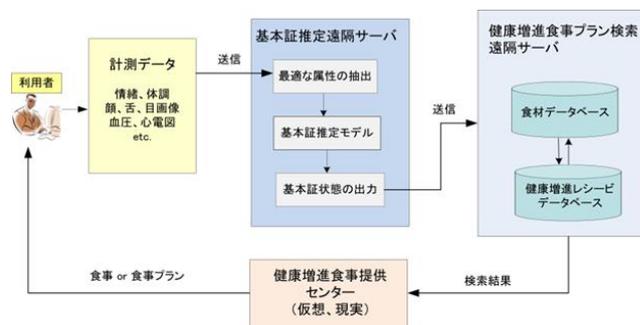


図1：研究流れ図

今までは特に下記の研究を行っている。

1) big sensor data から適切な特徴の抽出と統合

これまでの研究では、被験者 20 名の舌、目、顔などの外観画像及び、9 個の気分項目と WHO の健康測定による 25 個の体調項目のスコア、または、東洋医師 4 名による基本 15 証に関する診断データを学習データとして、基本証の推測システムを構築した。本研究では、これまで収集した上記のデータのほかに、血圧、脈拍、姿勢などの時系列センサーデータを加える。その上で、それらの異種センサーデータから健康状態や状態変化を反映する特徴を抽出し統合する方法を検討する。

2) 各状態間の相関性の分析

既存のプロトタイプシステムを利用し、first-person 評価による気分体調データと third-person 評価による基本証データとの相関性を分析する。その上で、Rich sensing を用いた日常健康状態の新たな表現方法を検討する。

3) 学習データの信頼度の測定

状態に関わる評価が人による評価であることから、多数の評価データの収集が困難であるうえに、それらのデータには曖昧性がある。そのため、学習データとする評価

データの質を判定する手法を検討し、状態推定の信頼度がある水準以上に保たれる方法を確立する。

4) 動的な状態感知と推定モデルの構築

利用者のニーズや使用状態によりセンサーのデータが欠けることがある。その影響を克服するために、状態感知モデルを動的に構築する必要がある。そのため、autonomous system の構築に関わる self-adaptation や node ensembles などの研究や deep learning などの技術を調査する。その上で、センサーの有無や変化に対処できる感知と推定モデルを動的に構築する手法を検討する。センサーデータの歪みにロバスト性がある状態推論アルゴリズムを開発し、有効性を検証する。

5) 学習データの収集とその充足度の判定

健康状態の感知と推定モデルを構築し検証するための学習データを用意する必要がある。本年度では、1)に述べた外観画像以外に血圧や脈拍などの項目を加え、学習データセットとして再構築する。20代から50代までの被験者約20名を対象とし、それぞれ半年間にわたり、少なくとも2週間1回の頻度でデータを収集する。被験者による気分と体調の評価データと東洋医師4名が診断する基本15証の評価データを合わせて、新異種学習データセットを構成する。収集されたデータのコストパフォーマンスを高めるために、モデルの構築に必要とされる学習データの件数を判定する基準について検証する。

3 これまで得られた研究の成果

本研究においては、提案された方法の有効性を検証するために、実世界のデータセットを用い、東洋医学の基本13証を推定するモデルを構築した。その訓練用なデータセットでは、さまざまな個人の健康状態に関するマルチモーダルセンサーデータが含まれている。これらのデータには、測定された身体状態のスコア、および一人称アンケートによって得られた主観的情報の報告が含まれる。さらに、個体の舌、顔、および目の画像から特徴が抽出される。各サンプルに対応するラベルデータは、4人の東洋医師が提供されたサンプルを診断した上で与えた基本13証に対するスコアである。一人称アンケートのデータ項目には、世界保健機関(WHO)が提案した9つの感情と健康状態に関連する13個の身体状態が含まれている。項目のスコアの範囲は0から5までである。顔および舌の画像から抽出された特徴が図2に示されている。

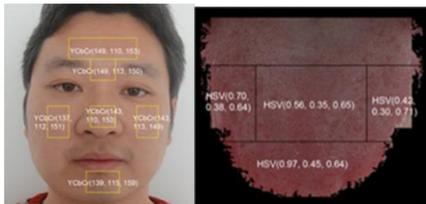


図2：抽出された特徴の例

トレーニングデータセットには、感情、身体状態、目、舌、顔の5つのモードがある。特徴の総数は71である。データセット中の32人からの150のインスタンスがあり、それぞれ71個の特徴と、東洋医4人に与えられた13

個の基本証スコアである。

定義された QoSTD がクラスの予測スコアの重みとして利用し、主観的クラスの分類性能を高めることを検証するために、二つの分類モデルが上記のトレーニングデータセットによって構築された。一つは matlab の fitsvm 関数を用い学習された support vector machine(SVM) 分類モデルで、もう一つが matlab の fitcdiscr 関数を用い学習された discrimination analysis(DA) 分類モデルである。

上記のバイナリ分類モデルに基づいて、インスタンスの各基本証に対するスコアは、MATLAB の predict 関数を利用して計算される。次に、QoSTD がスコアの weight として利用され、上記のスコアを調整する上でインスタンスの新たな尤度を計算する。

表1と表2はスコアを調整する前後のインスタンスに対する観測値とその尤度の一致度の結果を示している。表1は東洋医1が与えた13個基本証のラベルデータによって得た結果である。表2は東洋医3が与えたラベルデータによって得た結果である。

表1：東洋医1のラベルデータによる結果

TMCD1	DA				SVM		
	Zheng	QoSTD	Con_ori	Con	increased rate	Con_ori	Con
1	0.52	0.16	0.39	1.47	0.46	0.49	0.06
2	0.62	0.82	0.83	0.02	0.80	0.79	0.00
3	0.69	0.84	0.85	0.02	0.81	0.81	0.00
4	0.78	0.82	0.83	0.00	0.82	0.82	0.00
5	0.52	0.10	0.16	0.50	0.45	0.45	0.00
6	0.61	0.63	0.60	-0.04	0.63	0.63	0.01
7	0.46	0.23	0.36	0.59	0.44	0.46	0.06
8	0.63	0.83	0.83	0.00	0.83	0.83	0.00
9	0.59	0.77	0.82	0.06	0.70	0.70	0.00
10	0.43	0.01	0.10	13.40	0.29	0.37	0.30
11	0.74	0.90	0.90	0.00	0.91	0.91	0.00
12	0.75	0.81	0.82	0.00	0.81	0.81	0.00
13	0.41	0.10	0.31	2.05	0.53	0.55	0.03

表2：東洋医3のラベルデータによる結果

TMCD3	DA				SVM		
	Zheng	QoSTD	Con_ori	Con	increased rate	Con_ori	Con
1							
2	0.36	0.12	0.39	2.27	0.49	0.50	0.04
3	0.50	0.31	0.32	0.04	0.66	0.66	0.00
4	0.35	0.19	0.33	0.77	0.45	0.50	0.10
5	0.44	0.31	0.39	0.23	0.39	0.42	0.07
6	0.45	0.15	0.24	0.66	0.27	0.29	0.07
7	0.53	0.59	0.67	0.12	0.62	0.62	0.00
8	0.53	0.67	0.45	-0.33	0.68	0.68	0.00
9	0.41	0.37	0.07	-0.81	0.42	0.47	0.12
10	0.49	0.00	0.02	106.00	0.36	0.38	0.06
11	0.53	0.77	0.79	0.02	0.69	0.69	0.00
12	0.41	0.24	0.46	0.92	0.53	0.55	0.05
13	0.42	0.28	0.38	0.39	0.50	0.52	0.05

そこでは、Con が QoSTD をクラスの予測スコアの重みとして利用した時に、インスタンスに対する観測値とその尤度の一致率であり、Con_ori が QoSTD をクラスの予測スコアの重みとして利用しなかった時に、インスタンスに対する観測値とその尤度の一致率である。increased_rate が Con と Con_ori との変化率である。DA

[ここに入力]

[ここに入力]

[ここに入力]

欄が DA 分類モデルを利用した結果である。SVM 欄が SVM 分類モデルを利用した結果である。表の結果により、Con の値が increased_rate によって、大体の場合に増加した。特に、QoSTD が 0.5 未満の場合には、ほとんどの基本証に対して、Con の増加が比較的顕著である。

従って、QoSTD がクラスの予測スコアの重みとして調整すると、分類器のパフォーマンスが確かに向上していると言える。特に、分類器が QoSTD の値が低いデータセットで訓練されている場合、どんな分類モデルを利用しても、分類器の性能が著しく高まる。

表 3 と表 4 は分類器が多数回で訓練されたときの結果である。

表 3：東洋医 1 のラベルデータによる結果

TMCD1	DA				SVM		
	Zheng	QoSTD	max	max-1th	max order	max	max-1th
1	0.524	0.391	0.000	1	0.553	0.067	4
2	0.623	0.835	0.005	2	0.798	0.004	9
3	0.685	0.850	0.000	1	0.815	0.003	2
4	0.777	0.827	0.000	2	0.818	0.000	1
5	0.521	0.156	0.000	1	0.500	0.053	3
6	0.610	0.668	0.065	2	0.634	0.002	2
7	0.464	0.361	0.000	1	0.577	0.113	7
8	0.633	0.833	0.001	2	0.834	0.000	1
9	0.594	0.819	0.004	2	0.700	0.004	2
10	0.429	0.098	0.000	1	0.577	0.206	10
11	0.743	0.903	0.000	1	0.909	0.000	1
12	0.750	0.818	0.001	2	0.812	0.000	1
13	0.410	0.310	0.004	2	0.616	0.070	10

表 4：東洋医 3 のラベルデータによる結果

TMCD3	DA				SVM		
	symptom	QoSTD	max	max-1th	max order	max	max-1th
1							
2	0.36	0.895	0.503	4	0.590	0.085	6
3	0.50	0.772	0.511	3	0.851	0.187	6
4	0.35	0.328	0.000	1	0.689	0.193	5
5	0.44	0.707	0.322	5	0.622	0.207	5
6	0.45	0.244	0.000	1	0.633	0.340	5
7	0.53	0.828	0.159	2	0.794	0.175	6
8	0.53	0.897	0.448	5	0.858	0.181	6
9	0.41	1.000	0.931	3	0.797	0.328	6
10	0.49	1.000	0.979	3	0.766	0.388	6
11	0.53	0.878	0.120	2	0.759	0.072	6
12	0.41	0.884	0.425	5	0.777	0.223	6
13	0.42	0.918	0.536	5	0.811	0.289	6

そこでは、max が Con の最大値で、max order が Con が最大値になるときの訓練回数である。max-1th が最大 Con と分類器が第一回目訓練されたときの Con との差である。表の結果により、東洋医師や使用された分類モデルがどのようなものであっても、max と max-1th の差が 0 以上であることが分かった。また、東洋医 3 の場合には、QoSTD が比較的に低いが、max-1th の値が比較的に大きい。したがって、訓練データセットから分類未定なサンプルを削除し、洗練された訓練データセットで分類器を再トレーニングすれば、分類器のパフォーマンスを確実に向上させることができる。特に、QoSTD が比較的低いには、有効性が著しい。但し、トレーニングデータセ

ットでは、対応する証が必要なパフォーマンスを達成できないこともある。そのような場合には、これらのクラスの分類器を構成することは放棄されるべきである。

また、Con が最大値になれる再トレーニングのラウンドは異なることが注意すべき点である。学習データを洗練され、連続して分類器を再トレーニングすることは、分類の性能を改善しなければならないとは言えない。上記の問題を解決する方法は、Con を最大にするこのラウンドで訓練された分類器を採用すると考えている。

全体として、我々は、分類モデルを訓練するために実世界のデータセットを使用した。この訓練データセットは、東洋医師によってラベル付けされた 13 個の基本証に関わるスコアである。実験では、QoSTD が、絶対的な ground-truth がないクラスを分類するための分類器の性能に関連していることを実証しました。QoSTD と Con の間には高い正の相関がある。インスタンスの予測値を調整するための重みとして QoSTD を導入することで、分類器のパフォーマンスが確実に向上した。

4 今後の具体的な展開

今後の課題として、感情や性格などに関わる他の学習データを活用し、本研究で提案した方法の有効性を検証する。