

データマイニングによる緑内障診断システムの設計

Design of Glaucoma Diagnosis System by Data Mining

西山裕之*1 平石広典*2 岩瀬愛子*3 溝口文雄*1*2
 Hiroyuki Nishiyama Hironori Hiraishi Aiko Iwase Fumio Mizoguchi

*1 東京理科大学 理工学部 *2 株式会社ウィズダムテック *3 多治見市民病院
 Tokyo University of Science WisdomTex Inc. Tajimi Municipal Hospital

We designed a glaucoma diagnosis system by data mining. In this research, we applied the Support Vector Machine (SVM) of data mining tool for learning sampling data of fundus images and creating the diagnosis rules for our system. Using this diagnosis rules, we can detect the glaucoma images more accurately than general ophthalmologists.

1. はじめに

2000年に岐阜県多治見市で大規模に実施された緑内障調査 [Iwase04] の結果、予想以上の確率で発病者が存在するとともに、これまで有効とされてきた眼圧検査だけでは識別不可能な、低眼圧緑内障が過半数を占めていることが確認された。また、緑内障は点眼薬により進行を抑える以外の対処法は存在しないため、早期発見することが急務となっている。以上から、眼底写真を用いた定期的な診断が必要となるが、緑内障専門医の不足等の理由から実施は困難な状況となっている。また、レーザースキャン等の最新機材を用いたより高精度な診断方法も存在するが、現存する医院数は限られており、多人数に対する定期健診に用いることは現実的ではない。

このような背景より、本研究では、撮影された眼底写真を用いて緑内障の検出を自動化するための診断システムの設計を行う。緑内障診断を自動化する研究は、これまでも複数行われているが [Weinreb02]、レーザースキャン等の特殊な装置を必要とするものや [Weinreb02]、診断性能が眼科医に及ばないなど、汎用化および実用化が困難であった。本システムでは、従来の眼底写真から画像の RGB データの抽出を行うとともに [Hiraishi06]、SVM (サポートベクターマシン) [Ruping00, Vapnik95] を用いて数値データ群に対する学習を行うことで、眼底写真が緑内障眼を検知する診断ルールと緑内障眼の中の異常部位を特定する診断ルールの作成を行った。その結果、訓練データとして計 156 枚の眼底写真を用いて作成した検知用診断ルールは、クロスバリデーションの結果、全体の精度 (Accuracy) は 85.5%、異常眼に対する検出精度 (Sensitivity) は 90.4%、そして、非緑内障眼に対する検出制度 (Specificity) は 76.4% が得られた。この結果は、これまでの緑内障診断システム [Weinreb02] に比べて良好な結果である他、実際の眼科医以上の検知結果となっている。

2. サポートベクターマシン SVM

サポートベクターマシン (SVM: Support Vector Machine) は、教師あり学習を用いる分類手法の一つであり、パターン認識や回帰分析に適用できる。SVM の学習アルゴリズムは最適化理論から導出されており、Vapnik らによって 95 年に導入された学習法 [Vapnik95] は、訓練データとして与えられた正例と負例から、それらを分離する超平面の計算を行う。このとき、カーネル関数とマージン最大化といった工夫を加えることにより、高い識別性能を発揮できる [Nakaba05]。特に、非線

形写像を用いて高次元空間に写像し、その高次元特徴空間で線形分離を行うためのカーネル関数を導入することで、パターンを有限もしくは無限次元の特徴空間へ写像し、特徴空間上で線形分離を行うことができる。以上の特徴から、SVM はテキストのカテゴリ化、画像認識、そしてバイオインフォマティクス領域など、様々な分野で応用されている。

3. 眼底写真からのパラメータの抽出

緑内障の特徴として、乳頭部分の円形が崩れて血管に頸れが生じる。この乳頭の変型および血管の頸れを乳頭陥凹の異常と呼ぶ。また、網膜の視神経線維束欠損として、血管の下に線維が薄くなっている部分が存在し欠損が生じている (図 1 左参照)。以上より、眼底画像から識別可能な部分として、乳頭陥凹の異常もしくは視神経線維束欠損が考えられる。本研究では、特徴的には現れやすく視覚的な識別がより困難な視神経線維束欠損を患部として識別し、画像解析および SVM 学習により可能にする。

画像解析は、サンプリング画像の特徴をよりよく反映させるために、画像のフィルタリングおよび分割を行う [Hiraishi06]。また、フィルタリングは画像の二値化を用いて血管部分の抽出および排除を行うことも可能である。分割は 36 セグメントの扇状に分割する (図 1 右参照)。なお、各眼底画像は撮影条件により明るさに影響が生じるため、本研究では標準化を行った RGB データを用いている。これにより、一つのサンプリング画像は、36 (セグメント) × 3 (RGB) = 108 個の数値データ群として表現される。

4. SVM による緑内障診断ルールの作成

本研究では、上記により得られた各画像の RGB データに対し、SVM を用いて学習を行った。なお、SVM には代表的に利用されている MySVM [Ruping00] を使用する。訓練データとしては、緑内障眼を検知する診断ルールの作成には、緑内障の 102 枚の眼底画像と、非緑内障の 54 枚の眼底画像を用い、患部 (視神経線維束欠損) の部位を特定するための診断ルールの作成には、緑内障眼の眼底写真のうち、緑内障の専門医が患部部位を特定した画像 12 枚を用いた。各画像からの RGB データ抽出には、血管をそのまま残したものと血管を除去したものの 2 種類のデータ群に対して、個別に学習を実施した。

ここで、SVM を用いた学習方法および結果は、使用するカーネル関数およびデータの階層 (Dimension) により異なる。本研究では、カーネル関数として、広く良好な汎化性能を示すことが報告されている [Nakaba05]、Gaussian カーネル (下式) を用い、36 階層の RGB データすべてを繋げて、一枚の眼底画像を 108 の階層データの集合として扱った。

連絡先: 西山裕之, 東京理科大学理工学部, 〒 278-8510 千葉県野田市山崎 2641, 電話 04-7124-1501, hiroyuki@rs.noda.tus.ac.jp

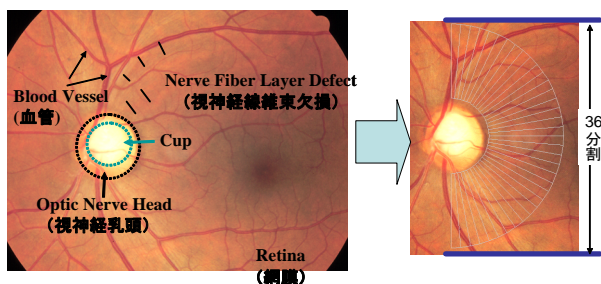


図 1: 眼底写真の各部名称・患部と眼底画像の 36 分割

	R	G	B	R	G	B	...	判定
↑ 156枚 ↓	0.9653	0.9728	0.9954	0.9513	0.9914	1.0294	...	1
	0.9627	1.0264	1.0571	0.9513	0.9914	1.0294	...	-1
	0.9627	1.0264	1.0571	0.9513	0.9914	1.0294	...	-1
	0.9627	1.0264	1.0571	0.9513	0.9914	1.0294	...	1
	← 108 + 1 階層のデータ →							

図 2: 緑内障を検知する診断ルールを作成する訓練データ

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2}\right)$$

4.1 緑内障を検知する診断ルールの作成

眼底画像ごとに、緑内障の場合は 1、非緑内障は-1 を専門医からの判定情報として付加した。その結果、使用するデータは図 2 のようになる。図 2 では、一行が 1 枚の眼底写真のデータ集合となる。これにより、眼底写真は 156 枚存在したため、データ集合は 156 行となる。

また、ここで、分類の粒度を制御するパラメータを調節する必要が生じる。そこで、パラメータを 1 ~ 10000 の範囲で学習実験を繰り返し行った。その結果、血管を削除しなかったデータ群、および削除したデータ群に対する SVM のクロスバリデーション結果で精度 (Accuracy) が最大となったものは、表 1 の様になった。

表 1: 緑内障を検知する診断ルールの精度

血管		Accuracy	Sensitivity	Specificity
あり	11.5	85.5 %	90.4 %	76.4 %
削除	9.4	83.3 %	87.3 %	75.9 %

4.2 部位を特定する診断ルールの作成

本研究では、緑内障を検知した後に、眼底写真内の患部を特定可能にするために、患部部位を特定する診断ルールの作成を行った。本学習には、緑内障専門医が図 1 の様に 36 分割された各セグメントに対し、患部 (視神経線維束欠損) の部位を特定した眼底写真 12 枚の RGB データ群を使用した。

本学習における訓練データは、36 分割された各セグメントごとに 1 件のデータとして扱わなければならないことから、図 3 の様に、セグメントの部位を表現するための RGB データを加えた後に、専門医の判定情報を付加している。これにより、眼底写真は 12 枚使用するため、データ集合は 12 枚 × 36 セグメントで 432 行となる。

上記の緑内障を検知する診断ルールの作成と同様に、パラメータを変更しながら SVM による学習を行った。なお、クロスバリデーションは、各眼底写真ごとに行った。その結果、血管を削除しなかったデータ群、および削除したデータ群に対する SVM のクロスバリデーション結果は、表 2 の様になった。

↑ 12枚 × 36セグメント ↓	R	G	B	R	G	B	...	部位R	部位G	部位B	判定
	0.9653	0.9728	0.9954	0.9513	0.9914	1.0294	...	0.9653	0.9728	0.9954	-1
	0.9653	0.9728	0.9954	0.9513	0.9914	1.0294	...	0.9513	0.9914	1.0294	1
	0.9653	0.9728	0.9954	0.9513	0.9914	1.0294	...	0.9627	1.0264	1.0571	1
	← 108 + 3+1 階層のデータ →										

図 3: 患部部位を特定する診断ルールを作成する訓練データ

表 2: 患部部位を特定する診断ルールの精度

血管		Accuracy	Sensitivity	Specificity
あり	11.0	66.0 %	45.9 %	82.6 %
削除	16.3	70.4 %	60.2 %	78.8 %

5. 考察

本実験の結果、緑内障を検知する診断ルール作成 (表 1 参照) においては、単なる眼底写真を用いているにも関わらず、レーザースキャナ等の特殊機器を用いた診断精度 [Weinreb02] と同程度以上の成果が得られた。また、Weinreb らが実験した実際の眼科医の精度 (Sensitivity: 88 %, Specificity: 59 %) [Weinreb02] よりも良好な結果が得られている。その他、画像解析による血管除去の必要性に関しては、表 1 より緑内障の検知に関しては血管が診断に関連しており、表 2 より患部部位の特定には、血管部分が阻害していることが確認できた。

6. おわりに

本研究では、眼底写真を画像解析して得られた画像データをサポートベクターマシンにより学習することで、緑内障の診断ルールの作成を行った。本ルールは一般的な眼科医以上の診断精度が得られた他、レーザースキャナ等の機器を用いた診断精度と同程度以上の結果となった。本診断ルールを用いることで、多人数に対する緑内障の診断システムを実現することが可能となる。

謝辞

東京医科大学の白土城照教授には研究のスタート時において眼底画像の判読方法を、また、東京理科大学の大和田勇人教授には、帰納学習のアプローチの試みが、研究のスタートになっている。グリッド研究環境については、サンマイクロシステムズの中西直之氏の支援を得た。これらの方々に感謝します。

参考文献

- [Hiraishi06] 平石広典, 西山裕之, 岩瀬愛子, 溝口文雄: 緑内障診断のための眼底画像解析システムの設計人工知能学会全国大会 (第 20 回) 論文集, 3A1-3, 2006.
- [Iwase04] Aiko Iwase, et al., The Prevalence of Primary Open-Angle Glaucoma in Japanese: The Tajimi Study *Ophthalmology*, Vol.111, No.9, pp.1641-1648, 2004.
- [Nakaba05] 中場優佑, 高橋由雅: 化学構造の TFS 表現を用いた SVM による薬物活性クラス分類, 人工知能学会全国大会 (第 19 回) 論文集, 1F1-04, 2005.
- [Ruping00] S. Ruping: mySVM-Manual, University of Dortmund, Lehrstuhl Informatik 8, 2000, <http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/>
- [Vapnik95] V. Vapnik: The Nature of Statistical Learning Theory, Springer-Verlag, 1995.
- [Weinreb02] R. N. Weinreb et al.: Comparing Machine Learning Classifiers for Diagnosing Glaucoma from Standard Automated Perimetry, *Investigative Ophthalmology and Visual Science*, Vol. 43, No. 1, pp.162-169, 2002.