カーネル法を用いたコンピュータ将棋の評価関数の学習

末 廣 大 貴† 畑 埜 晃 平† 坂 内 英 夫†
瀧 本 英 二† 竹 田 正 幸†

近年，コンピュータ将棋における評価関数は，機械学習を応用したパラメータの自動調整を行う手法が主流となっている．ただし，評価項目（特徴）は，作成者の考え，感覚に基づいて用意されることがほとんどである．本論文では，プロ棋士の棋譜を学習サンプルとして，カーネル法とサポートベクトルマシンを用いて学習を行う手法を提案する．カーネル法を用いることにより，作成者があらかじめ複雑な特徴を用意せずとも，局面を表現する単純な特徴のみから，特徴間の n 項関係などのより高次な特徴のが暗に生成され，その特徴空間で学習が行われる．複数の駒の位置関係の考慮が不可欠である囲いの評価実験を行い，カーネル法が有用性を示す結果を得た．また，本手法により得られた評価関数は，定跡などの明示的な知識を導入することなしに得られたにもかかわらず，特に序盤において，人間らしい局面評価を行うことを示す．

Learning Shogi Evaluation Functions using Kernel Methods

DAIKI SUEHIRO,† KOHEI HATANO,† HIDEO BANNAI,† EIJI TAKIMOTO†
and MASAYUKI TAKEDA†

Recently, automatic optimization of parameters by applying machine learning methods has become a mainstream approach for developing good evaluation functions in shogi. However, the features used in the evaluation functions are prepared by the developer, depending heavily on his/her knowledge and intuition. In this paper, we propose a method for learning evaluation functions from game records of professional players, using kernel methods and Support Vector Machines (SVMs). By using kernels, higher dimensional features such as n-ary relations between simple features can be considered implicitly, and various complex features can be considered without preparation. We apply our method on castle positions, which require consideration of relative positions of pieces, and show that the evaluation functions learned using kernels give better results. We also show that even without knowledge of standard moves, we were able to obtain human-like evaluation functions, especially in the opening.

1. はじめに

コンピュータ将棋における評価関数は，従来人間の手作業により調整されていたが，そこには大変な労力がかかるという問題と，作成者の棋力，感覚に依存してしまうという問題があった．しかし，保木によるBonanza Method と呼ばれる手法[1] の登場により，機械学習による調整の優秀性が示され，以来多くの強豪プログラムが機械学習による調整を行っている．Bonanza Method では，プロ棋士の棋譜を用いて，各局面においてプロ棋士の指した手を学習の教師とし，その手が最大の評価値を与えるように評価関数の学習を行う．

Bonanza Method の成功により，コンピュータ将棋は飛躍的な進歩を遂げたが，依然問題は残っている．局面評価にどのような評価項目（特徴）を用意するかについては，確立した手法がなく，いまだ作成者の棋力，感覚に依存している，という点である．現在評価項目として有力とされている特徴には，駒そのものの価値や，利きの数，駒同士の位置関係などから，個々の駒の働き具合などの細かいものまで，様々なものが考えられている．例えば駒同士の位置関係という特徴を，盤上の各駒同士の n 項関係と考えるとすると，どのような駒について，どのように n を設定するかは容易に決められるものではないし，多くの組み合わせを考えるほど計算時間は増大してしまう．本論文では，プロ棋士の棋譜を学習サンプルとして，カーネル法とサポートベクトルマシン（SVM）[2] を用いて学習を行う手法を提案する．SVM とは，2 値クラス分類に用いられる線形学習機械で，カーネル法と組み合わせることにより，極めて優れた性能を示すことが数多く報告されている．カーネル法とは，カーネルと呼ばれる非

† 九州大学大学院システム情報科学府
Graduete School/Faculty of Information Science and Electrical Engineering, The Kyushu University

線形関数を用いてデータをより高次元の特徴空間に暗に写像することで，非線形の概念の学習を実現する手法である．カーネル法を用いると，用いない場合（線形 SVM）にくらべ計算コストが多くかかってしまうというデメリットはあるものの，計算時間は特徴空間の次元の数によらない，というメリットがある．例えば n 次の多項式カーネルを用いることにより，先に述べた n 個の駒同士の位置関係を，それぞれ特徴空間上の次元とすることができ，これら n 項関係の全ての組み合わせを考慮して明示的に特徴ベクトルを構成する方法に比べ，大きく計算量を抑えることができる．また，複雑な特徴を考慮する労力から解放される，というメリットもある．本論文では，将棋の局面を表す上で，基本的で単純なの特徴のみを用い，線形 SVM で学習させたものと，カーネル法を用いて学習したものとを用意し，比較実験を行った．複数の駒の位置関係の考慮が不可欠である囲いの評価の実験では，カーネル法が有用性を示す結果を得た．また，定跡などの明示的な知識を導入することなしに得られたにもかかわらず，特に序盤において，人間らしい局面評価を行うことを示す．

2. 関 連 研 究

SVM やカーネル法を思考ゲームに応用した例として，麻雀[3] や将棋における詰みの予測[4] などがある．文献[3] では，木カーネルを用い，麻雀における特徴抽出とその重みの学習に成功している．文献[4] では，カーネル法こそコストの問題から使用していないが，SVM を利用して将棋の局面の特徴から詰みの有無を予測を行い，高い精度を残している．

3. SVM

事例空間を $X \subseteq R^n$，ラベル空間を $Y = \{-1,1\}$ とし，学習サンプルの集合を $S = ((\mathbf{x}_1,y_1),...,(\mathbf{x}_m,y_m)) \subseteq X \times Y$ とおく．$\mathbf{x}_i$ を事例，y_i を事例 $\mathbf{x}_i$ のラベルと呼ぶ．m は事例の数を表す．SVM は，入力としてサンプル S が与えられると，ソフトマージン最大化の原理に基づき，カーネルによって定義される特徴空間における線形関数 $f : X \to \mathbb{R}$ を求めるものである．通常，f は任意の事例 $\mathbf{x} \in X$ に対し，$f(\mathbf{x})$ の符号に基づいて 2 値クラス分類を行う仮説として用いられる．関数 f は，カーネル $K : X^2 \to \mathbb{R}$ を用いて

$$f(\mathbf{x}) = \sum_{i=1}^{m} y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^*$$

と表すことができる．ここで，α_i^*, $(i = 1,2,\ldots,m)$ は，以下の 2 次計画問題の最適解である．

maximise

$$W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j K(\mathbf{x_i}, \mathbf{x_j})$$

subject to

$$\sum_{i=1}^{m} y_i \alpha_i = 0, \quad C \geq \alpha_i \geq 0 (i = 1,\ldots,m)$$

b^* は $\alpha_i^* \neq 0, \alpha \neq C$ を満たす実数，C は適当なパラメータである．

4. 提 案 手 法

本研究では，事例ベクトル $\mathbf{x}$ の各成分を表す特徴（低次の特徴）として，局面を表現する上で，基本的，単純な特徴のみを用いた．以下では，事例ベクトルを局面ベクトルと呼ぶ．局面ベクトルの各成分の内訳は，盤上にある各駒（先手後手で分け，成駒を含めた 28 種類）の数，持ち駒（先手後手，合わせて 14 種類）とその数，盤上の各枡に位置している駒の種類，盤上の各枡に対する利きの有無とその駒の種類である．例えば初期局面であれば，先手の歩が盤上に 9 枚ある，先手の金が盤上に 2 枚ある，1 三に後手の歩がある，1 四に後手の歩が利いている，などがそれぞれその局面のベクトルの成分となる．局面ベクトルの次元数 n は，全てで 4058 個となった．多くの強豪プログラムでは，こうした低次の特徴の他に，作成者の感覚により有用と判断された複雑な特徴が多数追加され，膨大な次元数の局面ベクトルが用いられている．例えば，玉と金の位置関係や，玉の周囲の利きの数などがあげられる．本研究では，こうした特徴の多くが低次の特徴の d 項関係として表すことができる点に着目し，多項式カーネルを用いた SVM による評価関数の学習法を提案する．d 次の多項式カーネルは，

$$K(\mathbf{a}, \mathbf{b}) = (\mathbf{a} \cdot \mathbf{b} + 1)^d$$

で定義される関数で，対応する特徴空間 $\mathbb{R}^N$，$N = \binom{n+d}{d}$ の各次元が，高々d 次の単項式に対応する．すなわち，この特徴空間は，低次の特徴の d 項関係のうち，単項式として表すことができるものを全て特徴として含んでいる．また，カーネルの計算量は次数 d に依存しないことから，特徴空間内の次元数 N がどれだけ大きくとも，学習に要する計算量は変化しない．SVM の学習に際して，サンプルとしては，保木の手法などと同様，プロ棋士の棋譜を用いる．棋譜上のそれぞれの局面において，棋譜における次局面を正例，その他合法手により作成した局面全てを負例として，入力サンプルを作成する．このサンプルを SVM に与えて得られる関数を f とする．

対局を行う際は，f を局面の評価関数として用いる．すなわち，局面 $\mathbf{x}$ に対し，$f(\mathbf{x})$ が $\mathbf{x}$ の評価値となる．

5. 実　　験

5.1 学　　習

4 章に記した特徴と手順に基づき，SVM とカーネル法を用いて学習を行った．なお，SVM を実装するにあたり，既存のソフトウェアパッケージである SVM-light（`http://svmlight.joachims.org/`）を用いた．カーネル関数には 10 次の多項式カーネルを用いた．また，無作為に選んできたプロ棋士の棋譜 100 局分 2 セットから入力データサンプルを作成し，それぞれ線形 SVM を用いた学習，カーネルを用いた学習を行った．SVM-light は学習サンプルを 1 度にメモリに載せる必要があるため，今回は学習サンプルを 100 局分 (約 2GB) ほどまでに限定している．

5.2 対局による実験

それぞれ学習を行って得た評価関数 f をコンピュータ将棋に組み込み，4 章の方法で対局させた．ただし，min-max 探索は用いず，合法手後の局面でもっとも評価値の高い局面に導く手を次の一手として選択する．また，全てに 7 手詰みルーチンを組み込んだ．結果，先後入れ替えを含め，カーネルを用いたものが 8 勝 0 敗と，カーネル法を用いたものの全勝となった．特に，序盤の駒組みにおいて，カーネルを用いたものに学習の成果が強く見られた．また，人間や他コンピュータとの対局を観察しても，カーネルを用いたものの方は，手順などに問題はあるものの，駒得を積極的に行う，玉を囲う，飛車先の歩を突いていく，相手の駒を攻める，といった手が多く見られた．一方線形 SVM での学習の方は，玉を移動させるなどの駒組みを行う程度で，意味を持たない手が大半を占めていた．ただし，線形 SVM の場合は 100 局学習が数時間程度で完了するが，カーネルを用いた場合は 100 局分のデータで約 200 時間程度の計算時間を要した．

5.3 囲いの評価による実験

玉の囲いの評価が正しく行えるかどうかは駒の位置関係，組み合わせを考慮しているかが非常に重要であり，カーネルの効果を確かめる有効な手段と言える．今回は，美濃，穴熊，矢倉囲いの 3 つの囲いについて，それぞれ囲いを完成させる直前の局面を入力として用意し，囲いを完成させた局面が他の合法手後の局面に比べどう評価されているかを調べた．入力とする局面を図 1 ～ 図 3 に記す．また，評価関数 f には，線形 SVM で 100 局学習したもの，カーネルを用いて 100 局を学習したものの 1 つを用いた．結果は表 1 の通り

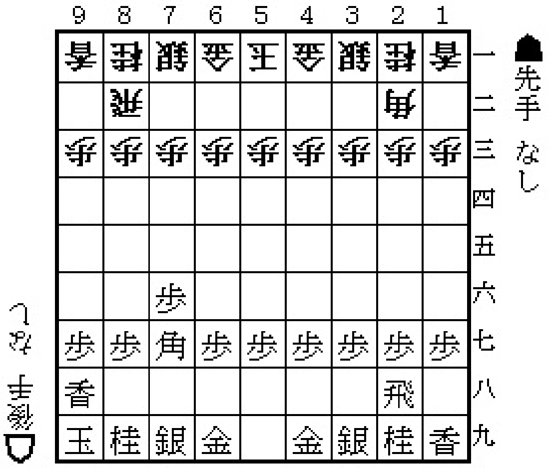

図 1　穴熊完成直前の局面（８八銀で完成）

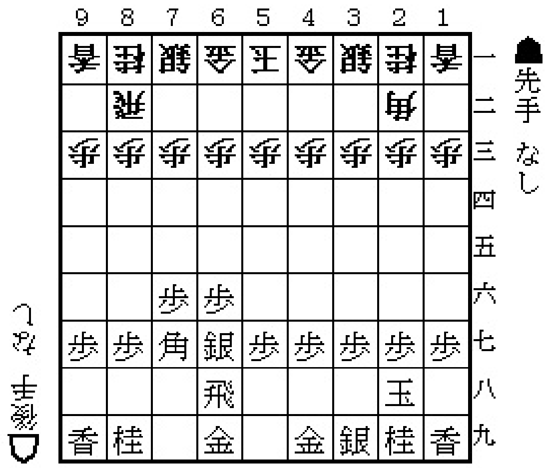

図 2　美濃完成直前の局面（３八銀で完成）

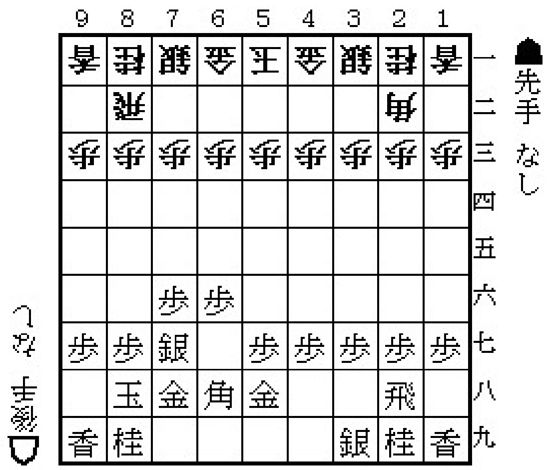

図 3　矢倉完成直前の局面（６七金右で完成）

表 1　囲いの評価

	カーネルと SVM	
	囲い完成	その他
局面 1(穴熊)	-0.99835753(1 位)	-1.0014502 ～ -1.0460587
局面 2(美濃)	-1.0046964(3 位)	-1.0032542 ～ -1.0259786
局面 3(矢倉)	-1.0111242(1 位)	-1.022269 ～ -1.0137082
	線形 SVM	
	囲い完成	その他
局面 1(穴熊)	-1.0000362(33 位)	-0.99982422 ～ -1.0001819
局面 2(美濃)	-1.0002418(24 位)	-1.0000873 ～ -1.0004555
局面 3(矢倉)	-1.0001571(7 位)	-1.0000775 ～ -1.0004269

になった．表中の値は f の値を示しており，この値が大きいほど，入力局面に対して高い評価をしていることになる．

カーネルを用いて学習を行った方は，囲いを完成させた局面に対して高い評価をしていることがわかる．一方線形 SVM を用いた学習の方は囲いを完成させた局面に対して特に高い評価値をつけるということはなかった．カーネルを用いることで，駒の n 項関係という次元を自動で作成し，学習していることを示すことができた．

5.4　駒組みの観察

図 4 ～図 6 は，本手法により作成されたプログラム（STR）が，人間との対局，および他のコンピュータ将棋ソフトと対戦した際に現れた局面である．なお，対局には，5.2 と同様の手法を用いている．図 4 は，第 1 著者との対局中に現れた局面で，居飛車急戦対四間飛車のごく自然な序盤である．図 5 は，第 20 回世界コンピュータ将棋選手権において，1 次予選で ponanza と対戦した際に出現した，横歩取り 8 五飛型の局面である．図 6 は，第 15 回コンピュータオリンピアード将棋部門で，BONANZA との 1 局において現れた局面である．BONANZA の四間飛車穴熊に対し，居飛車から美濃囲いへの駒組みを行っている．これらが示すように，わずか 100 局の棋譜データから，定跡等の知識を導入することなく得られた評価関数を用いて，人間らしい自然な駒組みを行えることは，驚くべきことである．

6.　まとめと今後の課題

本論文では，SVM とカーネルを用い，コンピュータ将棋の評価関数の学習法を提案し，その有効性を示した．特に，定跡や囲いの定形などの明示的な知識を与えずとも，100 局程度の棋譜からの学習により「人間らしい局面評価」を少なからず再現することができたことは，序盤の駒組みには難があると言われてきたコンピュータ将棋であるがゆえ，非常に大きな意味を持つと言える．

今後の課題としては，当然ながら，さらなる学習精度の向上，計算時間の高速化があげられる．そのためには，以下のような手法を取り入れることが考えられる．

- オンライン学習
  より精度の高い学習を行うためには，多くの学習サンプルを用いることが必要である．事実，多くの強豪プログラムは，数万局もの棋譜データから学習を行いっている．しかし，例えば約 5 万局から作成した学習サンプルは，約 700GB ものデータとなり，学習サンプルを一括して読み込む SVM-light を用いた実装では．メモリに載せることができず機能しない．実際，領域計算量は，サンプルの大きさの 2 乗に比例するため，学習サンプルが増大するにつれ，非常に大きな計算量を要してしまう．この問題を解決するためには，学習サンプルから局面ベクトルを読み込むたびに逐次パラメータ更新を行う，オンライン学習アルゴリズムを導入することが有効である．現在，代表的なオンライン学習アルゴリズムの 1 つである Pegasos[5] を用いた実装を行っており，5 万局サンプルから数日程度で学習を行うことに成功している．
- ランキング学習
  本来 SVM は 2 値分類を行う機械であり，仮説である評価関数の符号には意味がある（正ならプロ棋士が打ちそうな手，負ならそうでない手）ものの，その絶対値の大きさに理論的な意味を見出すことは困難である．従って，例えば合法手後の局面を全て負例と評価したとき，その中でもっとも評価値の高い局面が，最善の局面だという保証はない．そこで，全ての相対的な評価を行うことできるランキング学習[6)7)] の導入が有効であると考えられる．ランキング学習とは，正例（プロ棋士の手）に対する評価値が，負例（それ以外の手）の評価値よりも相対的に大きくなるようにパラメータ更新を行う学習法で，ゲームの評価関数の学習に適している．また，合法手を評価値の降順でソートすることにより，次の一手の第 1 候補手だけでなく，第 2，第 3 候補も提示することができる．
- min-max 探索を組み合わせた学習
  より正確な局面評価を行うためには，ある程度 min-max 探索を行うことが必要であろう．従って，学習のフェーズにおいても，サンプルに含まれる局面ベクトルをそのまま事例として用いるのでは

なく，min-max 探索を行って得られる数手先の局面を事例として用いることが考えられる．しかし，ここでの min-max 探索は，教師（プロ棋士）の読みによるものでなければならないが，そのようなデータは存在しない．従って，min-max 探索を組み込んだ場合の学習の目的関数として，3 章で記したような二次計画問題としての定式化ができそうもない．そこで，現在の評価関数を用いて行うという，標準的なヒューリスティックを用いることなどが考えられる．

参 考 文 献

1) 保木邦仁. 局面評価の学習を目指した探索結果の最適制御. 第 11 回ゲームプログラミングワークショップ, pp. 78–83, 2006.
2) John Shawe Taylor and Nello Cristianini, editors. *Suppor Vector Machines and other kernel based learning methods*. Cambridge University Press, 2000.
3) 三木理斗. 木カーネルを用いた SVM による麻雀打ち手の順位学習. 第 13 回ゲームプログラミングワークショップ, pp. 60–66, 2008.
4) 三輪誠, 横山大作, 近山隆. SVM を用いた詰みの予測とその応用. 第 9 回ゲームプログラミングワークショップ, 2004.
5) Shai Shalev-Shwartz, Yoram Singer, and Nathan Srebro. Pegasos:Primal Estimated sub-GrAdient SOlver for SVM. In *25th International Conference on Machine Learning(ICML)*, 2008.
6) Thorsten Joachims. Optimizing search engines using clickthrough data. pp. 133–142, 2002.
7) Eiji Takimoto Masayuki Takeda Jun-ich Moribe, KoheiHatano. Smooth Boostring for Margin-Based Ranking. In *Proceedings of the 19th international conference on Algorithmic Learning Theory*, pp. 227–239, 2008.

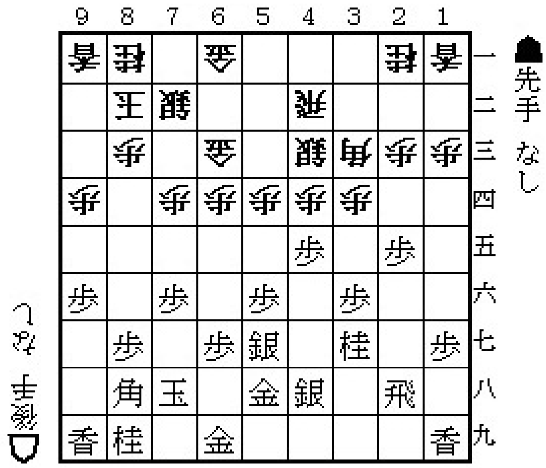

図 4 図は 33 手目先手 4 五歩まで（先手:STR）

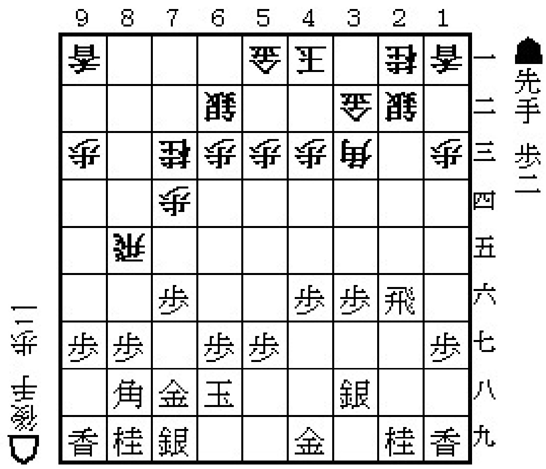

図 5 図は 30 手目後手 5 一金まで（後手:STR）

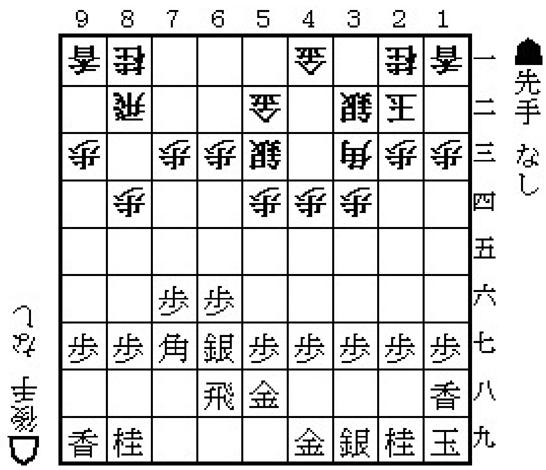

図 6 図は 24 手目後手 4 四歩まで（後手:STR）