

かかり受け構造を持つ文節に潜在変数が張り付く CRF モデルの学習不安定性

小早川 健
NHK 放送技術研究所

kobayakawa.t-ko@nhk.or.jp

1 はじめに

意見文を分類するタスクにおいて、文のかかり受け構造に着目した潜在変数モデル [1] が提唱されている。このモデルは、文の部分的な意味を担うための潜在変数を、文節かかり受け構造のノードに配置するもので、ノードに配置された変数が潜在的であることから、文の部分的な意味を局所的に決定することなく、曖昧性が保持されることを目指している。文全体の意味は、文に含まれる潜在変数の取りうる値を、すべての組み合わせについて足し上げることによって決定される。

このモデルは、自然言語の曖昧性を取り扱うためのものとして魅力的である。しかしながら、事例からのモデル学習における数値計算の安定性に欠けることが兼ねてから指摘されてきた。この不安定性は、大域的でない局所最適解をこのモデルが持つためだと考えられ、モデルパラメータの初期値を乱数によって試行する学習方法が取られてきた。

この研究では、この不安定性に着目して、モデル学習の数値計算実験と考察を行う。モデルパラメータの最適化法として、焼きなまし法という古典的な大域的最適化法を採用し、このモデルを詳細に調べた。その結果、

1. このモデルの持つ学習時の数値的不安定性は、モデルの持つ根元的な性質に由来していること
2. モデルの過度な自由度を制限するような拘束条件を付加することが効果的であること

が見出せたので報告する。この報告は、筆者が [2] において潜在変数間の拘束条件を導入する手法を提案したものについて、提案手法の挙動を詳細に考察した内容となっている。

2 タスク

意見文を種別に応じて分類するタスクを扱う。ブログなどの文に、構文・照応・評価情報を付与したコーパス (KNB コーパス)[3] を用いる。このコーパスに付与されている評価情報は、当為・要望・感情・批評・メリット・採否・出来事といった幅広い評価タイプのタグが付与されているが、このうち批評の評価タイプに着目し、批評 +/批評-の 2 値分類をこの研究の評価セットに用いる。

3 モデル

意見文の分類モデルは、中川ら [1] が提唱している Latent Dynamic CRF (図 1) を扱う。

CRF モデルのオリジナルは、

$$\log p_{\Lambda}(\mathbf{y}|\mathbf{x}) \propto \sum_{v \in V, k} \mu_k g_k(v, \mathbf{y}|v, \mathbf{x}) + \sum_{e \in E, k} \lambda_k f_k(e, \mathbf{y}|e, \mathbf{x}) \quad (1)$$

という形式の識別モデルであり、木構造の頂点集合 V の要素 v と辺集合 E の要素 e に素性 μ, λ が張り付くものである。分類ラベル s_0 に分類するモデル

$$\arg \max_{s_0 \in C} p_{\Lambda}(s_0|\mathbf{x}) \quad (2)$$

に一連の潜在変数 \mathbf{s} を含めると、

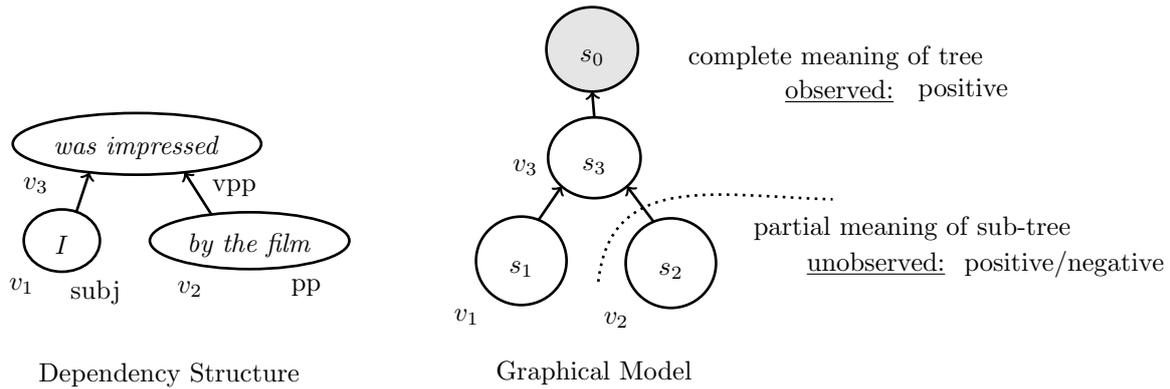
$$p_{\Lambda}(s_0|\mathbf{x}) = \sum_{\mathbf{s}} p_{\Lambda}(s_0, \mathbf{s}|\mathbf{x}) \quad (3)$$

の形で表されるので、入力文 \mathbf{x} が文節かかり受け構造に解析された木構造

$$\mathbf{x} \xrightarrow{\text{Dependency Parsing}} G(\mathbf{x}) = \{V(\mathbf{x}), E(\mathbf{x})\} \quad (4)$$

と融合する。

中川らの提唱する Latent Dynamic CRF モデルは、オリジナルの CRF モデルに対して、潜在変数が各ノード



eg. *I was impressed by the film.*

図1 LDCRF モデル

ドに張り付いたものとなっており、ノード数からなる一連の潜在変数を含むモデルとなっている。

このモデルに用いる素性は、図2に示すように、ポジ・ネガの評価極性を持つ表現の一覧である評価極性辞書、否定表現の一覧である反転語辞書、形態素の表層系であり、それぞれが、頂点 v や辺 e に張り付く。



図2 素性

モデルパラメータには、頂点素性が分類に寄与する場合の事後確率

$$\log p_{f_v}(s) = \mu_{f_v, s} \quad (5)$$

などが含まれ、モデルパラメータの集合 Λ は、正解事例の集合 \mathcal{D} から次式で学習される。

$$\mathbb{L}(\Lambda; \mathcal{D}) = \sum_{\mathbf{x} \in \mathcal{D}} \log \left(\sum_{\mathbf{s} \in C^{|\mathcal{V}(\mathbf{x})|}} p_{\Lambda}(s_0, \mathbf{s} | \mathbf{x}) \right) \quad (6)$$

筆者は、[2]において、潜在変数間の遷移ペナルティー

$$\mathbb{P}(\Lambda) = C_{\text{penalty}} \sum_{f_e \in \mathcal{E}(\mathcal{D})} \left(\sum_{s_1=s_2} (\log p_{f_e}(s_2|s_1) - \log C_{\text{same}})^2 + \sum_{s_1 \neq s_2} (\log p_{f_e}(s_2|s_1) - \log C_{\text{different}})^2 \right) \quad (7)$$

を提案し、このペナルティーを尤度関数に含めた学習による性能改善を確認している。ただし、このペナルティーが十分強い時にもモデルが正しい確率モデルとなるように、

$$C_{\text{same}} + (n-1)C_{\text{different}} = 1 \quad (8)$$

という条件を付加している。

式(7)の C_{penalty} は、この遷移ペナルティーの強さを制御するパラメータであり、次節において挙動を見ることにする。

4 大域的最適化による潜在変数モデル学習

潜在変数を含まないCRFモデルは凸性を有しており、パラメータ学習の初期値に関わらず、同一の最適解に収束する。一方で、ここで扱うような、潜在変数を含むモデルは、大域的でない局所最適解が存在する可能性がある。局所最適解の存在は、パラメータ学習を実施すればすぐにわかることで、このモデルには実際に局所最適解が存在する。

この研究では、大域的最適化として焼きなまし法 (Simulated Annealing) を採用する。焼きなまし法とは、モデルパラメータの変動を試行によって繰り返し、試行の採否を確率的に決定することが特徴である。試

行の採否を決定する閾値は、統計力学における状態遷移確率を基に算出され、温度の概念に相当するパラメータが0に向かうに従って、モデルパラメータが大域的最適解の付近に収束することが知られている。

モデルパラメータが大域的最適解の付近にあり、かつ、その大域的最適解が唯一である場合には、モデルパラメータ変動の試行の大多数は却下されることになる。逆に、モデルパラメータが最適解の付近にない場合や、付近に最適解が複数存在するときは、試行の採択率が一定の割合で存在し続ける。モデルパラメータが局所最適解付近にある場合には、大域的最適解付近にある場合と同様の挙動を示すが、確率的に局所最適解から脱出する可能性を秘めており、脱出後に逆戻りする可能性よりも脱出できる可能性のほうが高いことが、焼きなまし法の本質的な部分である。

Latent Dynamic CRF モデルを焼きなまし法で学習する場合の採択率の推移を図3に示す。縦軸は採択率で、横軸が計算の繰り返し回数である。ここでのパラメータ変動の試行は素性ごとに行われ、すべての素性について試行が試された時点で1回の計算繰り返し(iteration)とみなす。オリジナルの Latent Dynamic CRF モデル ($C_{\text{penalty}} = 0$ のモデル) は収束が悪く、採択率がなかなか減少しないことがわかる。この高止まりの原因は、最適解の縮退によるものと推測される。一方で、ペナルティ項を付加したもの ($C_{\text{penalty}} > 0$ のモデル) は、収束が早い。学習データの尤度の推移は、早晚100回程度の繰り返し回数で高止まるが、その後も採択率の減少は続いている。

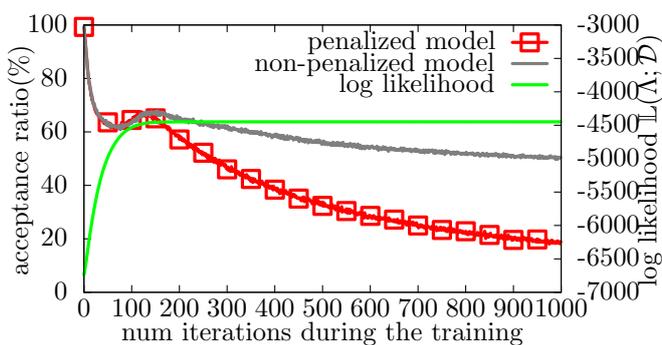


図3 乱数試行の採択率

この採択率の減少の原因が、オリジナルモデルにおける最適解の縮退が解けていることによるものと考えられたため、潜在変数間の取りうる値を拘束する項の強

弱による分類率の変化を調べた。図4には、横軸にペナルティ項の係数、縦軸に分類率が示されている。ペナルティ項が0の時に分類率と比較して、ペナルティ項が0でない場合には、分類率が改善されている。このペナルティ項には最適値が存在するはずであるが、どこかにピークを持つようには見受けられない。むしろ、0に近ければ近いだけよいように観察される。

これは、縮退を解くために必要なペナルティは弱いもので十分であることと合致する。つまり、同一の最適値に差が生じるためには、有限であればどんなに弱いペナルティであっても構わないということである。実験における数値計算に無限小という概念を持ち込むことは、丸め誤差(桁落ち)が生じるために難しい。しかし、図4の横軸に示してあるように、 $10^{-5} \sim 10$ の範囲でおしなべて改善効果が見られることは、このペナルティの項に役割が、縮退を解くためにのみ存在していることを示唆している。実際、ペナルティ項が10のように強い場合には、縮退を解くだけでなく、オリジナルモデルの性質を変えてしまうために、分類率が減少することが見て取れる。このことをはっきりさせるために、有限のペナルティが付加された場合の分類率を最小2乗法であてはめた直線を図に表示する。

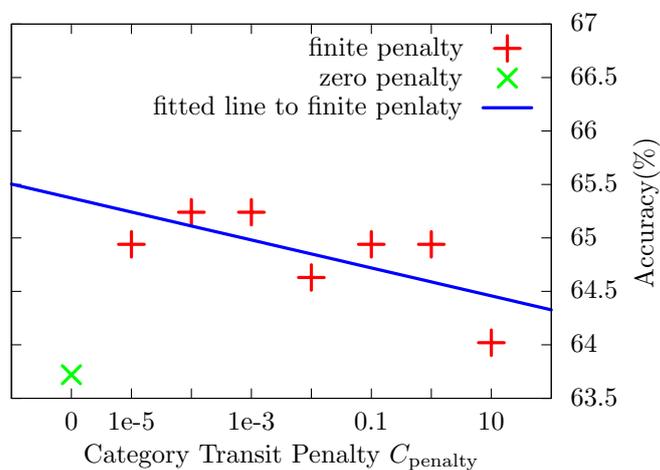


図4 弱いペナルティの効果

5 考察

対象とする現象をモデル化するにあたって、モデルに過度の自由度が含まれる場合には、縮退の問題が生じうる。このような場合の対処法として、正則化がよ

く知られている。しかしながら、正則化は、素性に冗長性がある場合に効果がある方法であって、潜在変数の冗長性に対して効果があるわけではない。正則化項をいくら強くしても分類率が向上しないことが、予備実験によって観測されていた。

また、学習によってパラメータが移動する割合も観測すると興味深い。2値分類の問題で、ペナルティ項が含まれないモデルで、最適化法の違いを比較したものが図5である。モデルパラメータの事後分布によって、学習空間上の位置を横軸に示してある。縦軸は、横軸のパラメータ値を持つ密度分布である。事後確率が同一になるように固定された初期値から出発して学習が終了する時点でのモデルパラメータの位置の分布が図示してある。分布の平均には、縦線が引かれている。学習が終了した時点で、最急降下法 (SD) で学習されたモデルでは、モデルパラメータが初期値の付近に停留するのに対して、焼きなまし法 (SA) で学習されたモデルでは、モデルパラメータが初期値から大きくずれた位置に移動している。

図5では、わかりやすさのために、すべての素性の分布ではなく、頂点素性に関する典型的な一部のモデルパラメータについてのみの分布が示されている。それは、評価極性辞書によって肯定的 (positive) の素性を持つモデルパラメータの肯定的な事後確率の分布、および、評価極性辞書によって否定的 (negative) の素性を持つモデルパラメータの否定的な事後確率の分布である。具体的に、横軸は、式 (5) にある $\mu_{f_v, s}$ の指数であり、縦軸は、その値を持つモデルパラメータがモデルパラメータ全数に占める割合 (確率密度) である。

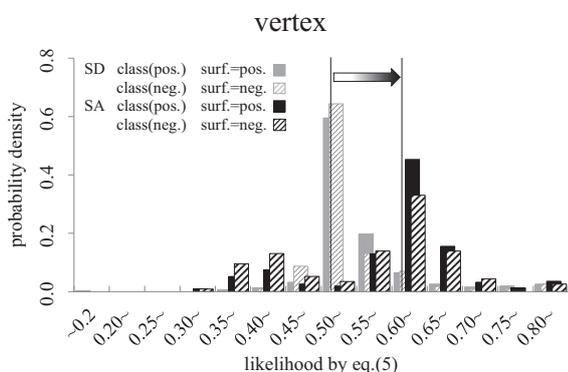


図5 頂点素性に対応するモデルパラメータの学習状況

この図からわかることは、拘束条件を付加しないオリジナルの Latent Dynamic CRF モデルは、最急降下

法 (SD) のようなアルゴリズムで学習する場合に、モデルパラメータのほとんどが初期値から移動しないで学習が終了してしまうことである。そのような状況下でも、焼きなまし法 (SA) では、モデルパラメータの初期値からの移動が見られることは非常に興味深い。これは、素性と潜在変数が関連している部分がモデルに含まれるために、わずかな学習データによる尤度差を生じていて、このようなわずかな尤度差に対して、焼きなまし法は頑健にモデル学習ができていたと考えられる。なお、遷移ペナルティを付加したモデルでは、最急降下法 (SD) で学習されたモデルにおいても高い分類率を示すことから、モデルパラメータは初期値からきちんと移動していることが推測される。

6 おわりに

かかり受け構造を持つ文節に潜在変数が張り付く CRF のモデルパラメータ学習の挙動について調べた。実験によると、潜在変数間の拘束条件がない場合に学習が不安定になる。これは、拘束条件を付加したモデルのパラメータ学習の挙動が安定的であることからわかることであるが、拘束条件の強さを変動させた実験から、不安定性が縮退に由来するものであると考えられる。

潜在変数が複数含まれるモデルにおいて、潜在変数の取りうる値の入れ替えについての対称性がある場合には、この研究が対象とした問題と同様の状況となるため、同様の学習不安定性を持つことが想像される。このような、他のモデルについての検討は、今後の課題としたい。

参考文献

- [1] 中川 哲治, 乾 健太郎, 黒橋 禎夫: 隠れ変数を持つ条件付き確率場による依存構造木の評価極性分類, 情報処理学会研究報告, Vol. 2009-NL-192, No. 10, pp.1-7, July 2009.
- [2] Latent Dynamic Model with Category Transition Constraint for Opinion Classification, 25th International Conference on Computational Linguistics (COLING 2014)
- [3] 橋本力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明. (2011). 構文・照応・評判情報つきプログコーパスの構築. 自然言語処理 Volume 18, Number 2, pp.175-201.