УДК 519.7

ПОДХОД К РЕШЕНИЮ ЗАДАЧ ОПТИМИЗАЦИИ С ПРЕЦЕДЕНТНОЙ НАЧАЛЬНОЙ ИНФОРМАЦИЕЙ

© А. С. Анафиев

Таврический национальный университет им. В.И. Вернадского
факультет математики и информатики
пр-т Вернадского, 4, г. Симферополь, 95007, Украина
e-mail: *anafiyev@gmail.com*

Abstract. The formulation of the optimization problem with precedent initial information is proposed. The main problems and tasks of the constructing reliable schemes for solving such optimization problems are highlighted. The approach for solving such problems based on the loss function is described. The example based on the metric classifiers is considered. The new class of the collective learning by precedent is introduced.

Введение

Рассмотрим задачу оптимизации в общем виде

$$\operatorname{extr} f(x) \,/\, x \in \Omega \subseteq X, \tag{1}$$

где $f : X \to Y$ — целевая функция, оптимальный (близкий к оптимальному) аргумент x^* которой требуется отыскать в ходе решения задачи, X — множество объектов, Y — множество значений (ответов) и Ω — множество (область) допустимых решений (ОДР), из которого выбирается оптимальное решение x^*.

Если при этом целевая функция и/или область допустимых решений (система ограничений) полностью не заданы, то мы имеем дело с задачей оптимизации с неполными данными (слабоопределенную задачу оптимизации). Неполная информация об элементах оптимизационной задачи может быть задана различными способами. Далее будут рассматриваться задачи, в которых информация о целевой функции и/или ограничениях представлена в виде прецедентов. Более подробно такие задачи рассмотрены в [1].

Сформулируем слабоопределенную задачу оптимизации с прецедентной начальной информацией в общем виде.

Пусть X — множество объектов, Y — множество значений целевой функции, W — множество допустимости. Рассмотрим задачу (1), информация о которой задана в виде набора прецедентов $X^\ell = \left\{(x_i, y_i, w_i)_{i=1}^{\ell}\right\}$, где $x_i \in X$, $y_i \in Y$ (если при этом значение целевой функции неизвестно, то ставится прочерк «$-$»); $w_i \in W$ — определяет степень принадлежности объекта x_i области допустимых объектов. Например, если $W = \{0, 1\}$, то w_i определяет принадлежность объекта x_i множеству допустимых объектов: 1 — принадлежит и 0 — не принадлежит; если $W = [0, 1]$, то w_i можно охарактеризовать как вероятность принадлежности объекта x_i множеству Ω.

Необходимо построить алгоритм, который, в некотором смысле, наилучшим образом[1] определяет множество оптимальных объектов (оптимальный объект) и оптимальное значение целевой функции задачи (1) или сводит задачу к известной задаче оптимизации с полностью определенными данными, допускающую эффективное решение.

Одним из подходов к решению задачи оптимизации с прецедентной начальной информацией, конечно же, является разделение исходной задачи на две подзадачи обучения по прецедентам: восстановление целевой функции (обычно, задача регрессии) и восстановление области допустимых решений (задача классификации), решение которых приводит к полностью определенной оптимизационной задаче, решив которую, можно получить окончательное решение.

Однако, при таком подходе теряется важная информация: например, при восстановлении области допустимых решений определяющую роль может играть информация о значениях целевой функции на объектах обучения, а при восстановлении целевой функции важным может оказаться знание того на сколько далеко от границы класса расположен объект, значение в котором определяется. Кроме этого, после восстановления целевой функции и области допустимых решений может получится задача оптимизации, которая не допускает эффективного решения, что является недопустимым при решении реальных практических задач.

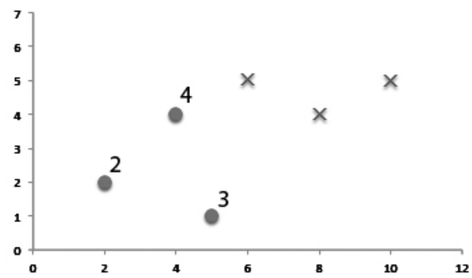
Таким образом, актуальной становится проблема построения схем решения оптимизационных задач с прецедентной начальной информацией, в которых процессы восстановления целевой функции и области допустимых решений выполняются ***совместно*** *друг с другом, «обмениваясь» между собой необходимой информацией. При построении таких схем важно учитывать, что решается именно задача оптимизации, а не две различные задачи обучения.*

Более того, мы получаем новый класс задач — класс *задач совместного обучения по прецедентам.* В нашем случае имеются две совместно решаемые задачи обучения по прецедентам: восстановление целевой функции и восстановления области допустимых решений, решение которых приводит к решению оптимизационной задачи. Следствием этого является необходимость изучения и разработки нового класса алгоритмов обучения, которые позволяют решать *совместно* целое *множество задач обучения* для достижения некоторого общего оптимального решения.

[1]Понятие «наилучшим образом» можно формализовать введением некоторого функционала качества.

Пример 1. Имеется следующая задача оптимизации (в данном случае, максимизации) по прецедентам: $X = \mathbb{R}^2$, $Y = \mathbb{R}$, $W = \{0, 1\}$. Обучающая выборка X^ℓ задана в виде обучающей таблицы:

x_1	x_2	y	w
2	2	2	1
5	1	3	1
4	4	4	1
6	5	-	0
8	4	-	0
10	5	-	0



Кружочками обозначены допустимые объекты, крестиками — недопустимые. Кроме этого объекты помечены значениями функции, если они известны.

Необходимо восстановить область допустимых решений и найти максимальное значение неизвестной целевой функции.

Как видно из рисунка, чем ближе мы приближаемся к воображаемой границе ОДР, тем выше значение функции на объекте. Очевидно, что данная информация, как уже отмечалось выше, не может не учитываться при решении данной задачи. И наоборот, при восстановлении целевой функции важным является расположение объектов из разных классов («ОДР» и «не ОДР») множества объектов X.

1. Функция потерь. Функционал качества

Аналогично задачам обучения по прецедентам можно определить функцию потерь и функционал качества и для задач оптимизации с прецедентной начальной информацией.

Функция потерь — это неотрицательная функция $\mathscr{L}(a, x)$, характеризующая величину ошибки алгоритма a на объекте x [2].

Функционал качества алгоритма a на выборке X^ℓ можно определить как суммарную потерю на всех объектах обучения:

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathscr{L}(a, x_i). \tag{2}$$

Обозначим через $a_f(x) \in Y$ ответ алгоритма a о значении целевой функции на объекте x, а через $a_\Omega(x) \in W$ — ответ алгоритма a о принадлежности объекта области допустимых решений. Тогда в качестве функции потерь, например, можно рассмотреть функцию $\mathscr{L}(a, x) = \alpha[a_\Omega(x_i) \neq y_i] + \beta\big(a_f(x_i) - y_i\big)^2$, где α, β — некоторые параметры алгоритма. Данная функция потерь совместно учитывает и правильность определения принадлежности к ОДР и качество восстановления целевой функции.

Одним из способов решения задач оптимизации с прецедентной начальной информацией является модификация функции потерь в сторону взаимного учета потерь при восстановлении целевой функции и области допустимых решений и последующим использованием алгоритмов обучения для решения задачи оптимизации по прецедентам.

Один из подобных методов рассмотрен в работе [3], в которой исходная задача оптимизации сводится, с помощью применения модифицированной функции потерь L_ε, к задаче регрессии, для решения которой предлагается использовать метод *Support Vector Regression* — SVR [4].

Используя огромный набор методов классификации и регрессии, можно получать большое количество подходов к решению задач оптимизации с неполными данными. Конечно же, такой подход порождает множество вопросов: «*какой набор методов машинного обучения и как необходимо использовать при решении той или иной конкретной практической задачи?*»; «*как вычислять адекватность полученных моделей и, следовательно, надежность полученных решений?*»; «*как правильно синтезировать различные алгоритмы обучения для построения схем решения задач оптимизации с прецедентной начальной информацией?*» и т.д.

2. Применение метрических алгоритмов классификации для решения задачи оптимизации с прецедентной начальной информацией

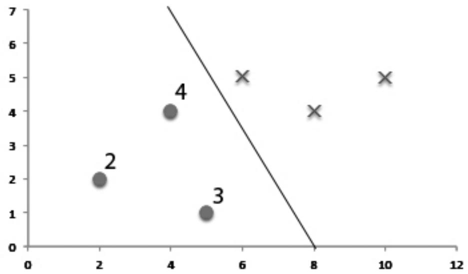
Как известно, метрические алгоритмы классификации основаны на гипотезе компактности, предполагающей, что схожие (по некоторой метрике) объекты расположены внутри одного класса, а объекты далекие по метрике в разных.

Обобщенный метрический классификатор можно записать в виде [2]:

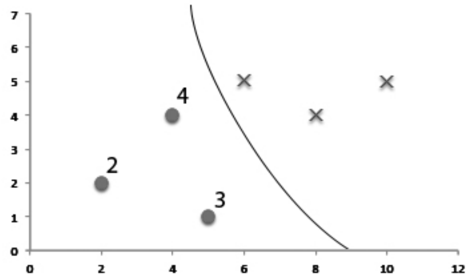
$$a(u; X^\ell) = \arg\max_{y \in Y} \Gamma_y(u, X^\ell); \qquad \Gamma_y(u, X^\ell) = \sum_{i=1}^{\ell} [y_u^{(i)} = y]\, w(i, u) \tag{3}$$

где u — классифицируемый объект, $y_u^{(i)}$ — метка класса i-го соседа объекта u, $\Gamma_y(u, X^\ell)$ — суммарный вес ближайших к u обучающих объектов из X^ℓ, $w(i, u)$ — весовая функция, оценивающая степень важности i-го соседа для классификации объекта u.

Рис. 1. Метод потенциальных функций для решения задач оптимизации



а) $h_i = 5$, $i = 1..6$.



б) Переменная ширина окна равная величине потенциала, $h_i = y_i$.

Рассматривая различные весовые функции, можно получать различные метрические алгоритмы классификации, а следовательно и различные методы решения задач оптимизации с прецедентной начальной информацией.

- Для метода ближайших соседей весовая функция имеет вид $w(i, u) = [i \leqslant k]$. Модернизируем ее для задачи оптимизации. Для этого будем учитывать не только номер соседа, но и значение целевой функции в нем: $w(i, u) = [i \leqslant k]K\big(y_u^{(i)}\big)$, где $K(z)$ — некоторая функция, характеризующая степень важности i-го соседа в зависимости от значения целевой функции.
- Для метода парзеновского окна и метода потенциальных функций достаточно переопределить функцию расстояния между обучающими объектами, так, чтобы она учитывала не только координаты объектов, но и значения целевой функции в них. Например, для метода потенциальных функций

$$a(u, X^\ell) = \arg\max_{y \in Y} \sum_{i=1}^{\ell} \gamma_i \, K\left(\frac{\rho(x_i, u)}{h_i}\right), \quad \gamma_i \geqslant 0,\ h_i \geqslant 0,$$

в качестве значений потенциалов γ_i можно рассматривать значения (или некоторую функцию от них) целевой функции на объекте, а если взять $W = \{-1, +1\}$, то принадлежность области допустимых решений можно рассматривать как знак заряда (см. рисунок 1).

Заключение

Сформулирована задача оптимизации с прецедентной начальной информацией в общем виде. Выделено направление применения широкого класса методов машинного обучения для решения задач оптимизации по прецедентам. Очерчены основные проблемы и задачи получения схем построения адекватных, согласованных с начальной информацией, моделей задач оптимизации с прецедентной начальной информацией.

Выделен новый класс задач обучения — класс совместного обучения по прецедентам.

Список литературы

1. Анафиев А. С. Оптимизационные модели с прецедентной начальной информацией / А. С. Анафиев, В. Ф. Блыщик // Таврический вестник информатики и математики. — Симферополь: КНЦ НАНУ, 2011. — №2. — С. 51–57.
2. Воронцов К. В. Математические методы обучения по прецедентам. Курс лекций по машинному обучению. [Электронный ресурс] / К. В. Воронцов. — 2011. — 141 с. Режим доступа к курсу: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf
3. Таратынова Н. Ю. Построение оптимизационной модели по прецедентной начальной информации как задача нелинейной регрессии / Н. Ю. Таратынова // Искусственный интеллект. — Донецк: Институт проблем искусственного интеллекта, 2006. — №2. — С. 238–241.
4. Smola A. J. Tutorial on Support Vector Regression / Smola A. J., Scholkopf B. A. // Statistics and Computing, 2004. — Vol. 14. — P. 199–222.

Статья поступила в редакцию 02.06.2012