ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0028

Emotional Classification Model based on Three Interval Decisions

Zhiwei Yang, Yuqing Sun, Xinru Li, and Xiaopei Li

School of North China University of Science and Technology, Hebei 63210, China

Abstract

This paper mainly applies three interval decision models to conduct emotional analysis of microblog comments.First, the emotion classification model based on the three-way interval decisions is constructed, based on the pre-processing of the raw data, we construct the formal background of 0-1, obtain the connotation and extension of the interval concept, divide the positive, negative and boundary fields according to the three decision theories, and establish the three-way decision classification rules. Secondly, the collected microblog comments are divided into training set and test set. Through pre-processing operation and using the conditional feature attributes, divide the comment users into three domains, calculate the loss value of the three domains, get the interval concept table as a classification model, substitute the test set as a new object into the model, and find the lowest cost in the model.Finally, the classification accuracy based on the interval three-way decisions is obtained.

Keywords

Interval Three Decision-Making, Formal Background, Interval Concept.

1. Introduction

With the increase of microblog users and the improvement of people's attention to the generated text information, the emotional analysis of applying various algorithms and models to microblogs and comments has gradually become a hot topic.The model developed in this paper aims to get good classification results by constantly substituting new text into the classification models to realize the emotional classification of microblog comments.In this paper, we apply three interval decisions to the emotional classification of microblog comments in order to obtain an algorithmic model with real-time dynamics.An emotion classification model based on interval three decision algorithm.The key steps include: the construction of the formal background, the formation of the interval concepts, the division of the domains and the calculation of the loss values, and the formulation of the classification standards.

2. Interval Three-Way Decision Algorithm

2.1 Three Decisions

Let u decision scheme set, c is a finite condition set, the goal of the decision is to make the corresponding decision for each entity according to a given condition, then give the basis for the decision, and different evaluation functions can be constructed according to different kinds of conditions.Given an entity, its evaluation function value is called decision state values, and the value domain of the evaluation function is generally a set of offsets, interval values, real numbers, lattices, and integers.Its construction can be based on cost, error, risk, profit and satisfaction.

The core idea of the three decisions is to introduce two threshold values and on the entity evaluation function, and to construct the three required domains and make the corresponding decisions:

(1) An entity is accepted when the decision status value is greater than or equal;

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0028

(2) An entity is rejected when the decision status value is less than or equal;

(3) When the entity's decision status value is between and, neither accept nor reject the entity, make a third decision: no commitment decision, also called delayed decision, requires more information or further observation.

Three decisions can be described as follows: for a finite set of non-empty entities and a finite condition set.It will be divided into three pairwise disjoint domains, called the positive, negative, and boundary domains, namely, recorded as,, and, respectively.At the same time, there are two states: meet and do not meet the given conditions, expressed as sum respectively.

An equivalence relation on the definition, if any, and the equivalent, is noted as.Equivalent classes consists of all objects that satisfy the equivalence relationship.The evaluation function represents the probability of dividing objects in an equivalent class into a set, and a pair of thresholds can divide the complete set into positive, negative, and boundary domains.The performance of the three decisions depends on the evaluation function and the threshold value selected, and the performance of each model needs to be analyzed by different metric measures.

2.2 Interval Three Decision-Making

The theoretical domain can be divided by the three-way decision concept and the upper and lower boundary extension into three regions[1]:

$$POS_{\alpha}^{\beta}(Y) = M^{\beta} = \{y \in u||f(x) \cap W|/|W| \geq \beta\} \tag{1}$$

$$BND_{\alpha}^{\beta}(Y) = M^{\beta} - M^{\alpha} = \{y \in u|\alpha|f(x) \cap W|/|W| < \beta\} \tag{2}$$

$$NEG_{\alpha}^{\beta}(Y) = u - M^{\alpha} = \{y \in u||f(x) \cap W|/|W| < \alpha\} \tag{3}$$

If $y \in POS_{\alpha}^{\beta}(Y)$, execute Accept for y, if $y \in NEG_{\alpha}^{\beta}(Y)$, reject for y, or No Commitment is executed.For the new entity y, the decision obtained from the three decision concepts of the interval is composed of the decision action b_i and the corresponding decision loss R, recorded as $j = (b_i, R(b_i))$, and the three decision space of the constituent interval is expressed as$\widetilde{l_{\alpha}^{\beta}}(u, c \cup d, R)$.

The three interval decisions essentially complete the clustering, and obtain the lattice structure through the relationship between the connotations in the concept to form the decision space.Father-son relationship to decide the subsequent decision to reduce the loss.

3. Emotion Classification Model based on Interval Three-Way Decision Algorithms

3.1 Interval Three Decision-Making

The main feature attributes are selected from the pre-processed[2] Weibo comments data as the conditional attribute data , and the formal background form is shown in Table 1:

The main feature attributes are selected from the pre-processed Weibo comments data as the conditional attribute data $c = \{c1, c2, c3 \cdots, cn\}$, and the formal background form is shown in Table 1:

In them, $id = \{1,2,3, \cdots, n\}$ is a document / text entry and c_i is the feature attribute. Finally, three decisions are made on the data, and the decision attribute set is: $D = \{d1, d2, d3 \cdots dn\}$, $d1, d2, d3 \cdots dn$ represents the category to which the text belongs.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0028

Table 1. Form background

id	c1	c2	c3	c4	……	cn	D
1	0	1	0	*	……	*	d1
2	*	*	*	*	……	*	d2
3	*	*	*	*	……	*	d3
…							…
n							d1

3.2 Build the Interval Concept

1)According to the ICAICL algorithm [3], we first calculate the attribute set power set $P(B)$, determine the connotation of the concepts, and generate the initialized concept set H.

2)The values of the threshold values α and β are given, and the junction corresponding to W is $H = (M^{\alpha}, M^{\beta}, W)$.For W scanning the connotation of each object, the text feature attribute. If the condition $W_i \subseteq W$ and $|W_i|/|W| \geq \alpha$ are met, the object d_i is incorporated into the M^{α} upper bound epitaxial α of H,similarly, if the condition $W_i \subseteq W$ and$|W_i|/|W| \geq \beta$ is satisfied, the object d_j is incorporated into the M^{β}-upper bound epitaxial β of H.

3)From the above parameters and rules, the interval form and concept table are obtained.

3.3 Domain Division

Suppose a formal background $(u, c\cup d, R)$, $\widetilde{H} = (M^{\alpha}, M^{\beta}, W; R(b_P), R(b_B), R(b_N))$,as interval three decision concepts.

The upper and lower bound extension can divide the theoretical domain u into three regions:

$$POS_{\alpha}^{\beta}(Y) = M^{\beta} = \{y \in u \,|\, |f(x) \cap W|/|W| \geq \beta\} \tag{4}$$

$$BND_{\alpha}^{\beta}(Y) = M^{\alpha} - M^{\beta} = \{y \in u \,|\, \alpha \leq |f(x) \cap W|/|W| < \beta\} \tag{5}$$

$$NEG_{\alpha}^{\beta}(Y) = u - M^{\alpha} = \{y \in u \,|\, |f(x) \cap W|/|W| < \alpha\} \tag{6}$$

The division and decision of the classic three decision domains.

3.4 Decision loss function

Set in the state set $S = \{Y, \overline{Y}\}$, represent the object set x and excluded set x in the y theory domain x, respectively and the not included set Y.Action set $B = \{b_P, b_N, b_B\}$, corresponding to the decision-making action: acceptance, rejection, and delay.The decision cost matrix is shown in Table 2:

In Table 2, the first column corresponds to the loss function values for decision acceptance, rejection, and noncommitment when y truly belongs to Y;similarly, the second column represents the loss value corresponding to the three decisions when y truly belongs to $\overline{Y}$.

The following are expressions for the loss function under three decision actions: acceptance, rejection, and noncommitment:

$$R(b_P) = \lambda_{PP}(|Y \cap M^{\beta}|/|M^{\beta}|) + \lambda_{PN}(|\overline{Y} \cap M^{\beta}|/|M^{\beta}|) \tag{7}$$

$$R(b_N) = \lambda_{NP}(|Y \cap (u - M^{\alpha})|/|u - M^{\alpha}|) + \lambda_{NN}(|\overline{Y} \cap (u - M^{\alpha})|/|u - M^{\alpha}|) \quad (8)$$

$$R(a_b) = \lambda_{BP}(|Y \cap (M^{\alpha} - M^{\beta})|/|M^{\alpha} - M^{\beta}|) + \lambda_{BN}(|\overline{Y} \cap (M^{\alpha} - M^{\beta})|/|M^{\alpha} - M^{\beta}|) \quad (9)$$

Table 2. Decision cost matrix

Decision action	The Objective state of an entity	
	Belong to *Y*	Do not belong to *Y*
Accept	λ_{PP}	λ_{PN}
Refuse	λ_{NP}	λ_{NN}
Delay	λ_{BP}	λ_{BN}

3.5 Classification Rules based on Three-Way Interval Decisions

Set the formal background $(u, c \cup d, R)$, the three decision concept table of the resulting interval is $\widetilde{H_{\alpha}^{\beta}}(u, c \cup d, R)$.

If v such concepts exist in the interval concept table, $\widetilde{H_{\alpha}^{\beta}}(u, c \cup d, R)$ "accepted" decisions are obtained,that is: $E_1, E_2, \cdots, E_v$., from these decisions constitute the v-dimensional decision space $E = \{E_1, E_2, \cdots, E_v\}$ The $E_j \in E$ is assumed to arise by the interval notion $\widetilde{H} = (M^{\alpha}, M^{\beta}, B; R(b_P), R(b_B), R(b_N),)$. For a new object y, the set of conditional properties is known to be $B(B \subseteq c)$.In the interval concept table $\widetilde{H}$, the concept of the conditional attribute is B in the search connotation. There are several following situations:

1)If a unique decision E_k meets the:

$$R_{Pk} = \min(R_{P1}, R_{P2}, \cdots, R_{Pn}) \quad (10)$$

The acceptance decision with the least loss is taken as the acceptance decision, which as the final decision on the new object y.

2)When there are multiple E satisfications (8), the properties of the new object y are further measured until a unique decision that can satisfy the condition.

For example, during the decision process, the decisions obtained by the interval three decision concepts $\widetilde{G}_m$ and $\widetilde{G}_n$ are E_m and E_n, respectively, while satisfied:

$$R_{Pm} = R_{Pn} = \min(R_{P1}, R_{P2}, \cdots, R_{Pn}) \quad (11)$$

Search for subconcepts $\widetilde{H}_m$ and $\widetilde{H}_n$ that add measurement properties according to the connotation w of these subconcepts to continue finding acceptance decisions with minimal loss.

4. Emotional Classification of Weibo Comments based on Three Interval Decisions

If you follow the “checklist” your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

4.1 Construction of the Formal Background

For the comments under this topic, we can take some of the following characteristic words as conditional attributes: ['no', 'want,', 'willing', 'school', 'see', 'efficiency', 'flower bai', 'money', 'senior three', 'graduation', 'home', 'incense'], that is $C = \{c1, c2, c3 \cdots, c12\}$.

The data were divided into training and test sets. For the training set, traverse each object in the training set, each comment, and store the background in the conditional attribute, such as "I want to go out, start school soon", and the formal background will be $[0,1,0,1,0,0,0,0,0,0,0,0,'1']$.The conditional attribute for this comment is $['c2','c4']$.Of these, '1' is obtained by the label of that comment, for the decision attribute of the comment.

4.2 Build a Three-Way Decision Concept Table in the Interval

First, we need to calculate the attribute set power set PA, which have 2^{12} power sets for this experiment due to the inclusion of 12 conditional properties,and there are $(2^{12} - 2)$ non-empty true subsets, and because the comments are divided into positive and negative, namely: '1' (positive) and '0' (negative), the final concept set of initialization should have $(2^{12} - 2) \times 2$.

Conduct circular traversal based on the obtained initialization concept set, combine the upper and lower epitaxial, connotation and 3.2.1 model building division principles, give parameters α,β and the value of the parameter λ_{PP},λ_{NP},λ_{BP},λ_{PN},λ_{NN},λ_{BN} for calculating the cost loss and calculate the objects contained in the upper and lower epitaxy and divide the object set u into accepted, reject, and uncommitted domains,that is $POS(y)$,$NEG(y)$and$BND(y)$.Another cost loss function computes the loss value of the three decisions:a_P, a_N, a_B.

Finally, the upper and lower bound epitaxial, connotation, acceptance loss, noncommitment loss, rejection loss are stored as hexples, and the classification model is obtained.

4.3 Classify

Consider each comment in the test set as something new to the model. If the new object has a conditional property: $['c1','c4']$, find the concept of the conditional property $['c1','c4']$ in all connotations in the model.

(1)If the concept has only one acceptance decision, then take the decision attribute of the concept as the determined category. The eligible concept is:

$$g1 = \left[[1,2,3,5],[2,4,5,6],['c1','c4','1'],0.5,3.2,6.4\right]$$

At this time, the category is divided into '1', which is judged as a positive emotion.

(2)If there is more than one accepted decision in these concepts, the decision attributes corresponding to the cost minimum of all accepted decisions are taken as categories. The concepts of conformity are:

$$g1 = \left[[1,2,3,5],[2,4,5,6],['c1','c4','1'],0.5,3.2,6.4\right]$$

$$g2 = \left[[1,2,3,9],[2,4,5,6,7],['c1','c4','c5','0'],0.3,4.3,5.6\right]$$

The decision should take $\min\{0.5,0.3\}$, that is, the acceptance decision of the concept $g2$ and divide the category into '0', which is judged as a negative emotion.

The final experiment showed the classification accuracy of 0.783.

5. Conclusion

This paper details the emotion classification process of microblog comments on the three interval decision models, describes the process of model construction in detail, and obtains the final experimental results.

References

[1] Wang Liya, Zhang Chunying, and Liu Baoxiang. Three decision dynamic strategy regulation models based on the interval conceptual lattice [J]. Computer Engineering and Application, 2016,52 (24): 80-84 + 101.

[2] Wang Tianji. Chinese microblog article based on three decisions [D]. Hunan University of Technology, 2018.

[3] Wang Liya.Research and application of efficient lattice algorithm [D].North China University of Science and Technology, 2015.