International Journal of Parallel, Emergent and Distributed Systems

ISSN: 1744-5760 (Print) 1744-5779 (Online) Journal homepage: http://www.tandfonline.com/loi/gpaa20

Blood typing for families: a novel hybrid human–computer application

Wenjun Jiang, Jie Wu, Guojun Wang & Huanyang Zheng

To cite this article: Wenjun Jiang, Jie Wu, Guojun Wang & Huanyang Zheng (2015): Blood typing for families: a novel hybrid human–computer application, International Journal of Parallel, Emergent and Distributed Systems, DOI: 10.1080/17445760.2015.1055268

To link to this article: http://dx.doi.org/10.1080/17445760.2015.1055268

Published online: 24 Jul 2015.

International Journal of Parallel, Emergent and Distributed Systems, 2015
http://dx.doi.org/10.1080/17445760.2015.1055268

Taylor & Francis
Taylor & Francis Group

Blood typing for families: a novel hybrid human–computer application

Wenjun Jiang[a]*, Jie Wu[b,1], Guojun Wang[c,2] and Huanyang Zheng[b,3]

[a]*School of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan Province, 410082, P.R. China;* [b]*Department of Computer and Information Sciences, Temple University, 1805 N. Broad Street, Philadelphia, PA 19122, USA;* [c]*School of Information Science and Engineering, Central South University, Changsha, Hunan Province, 410083, P.R. China*

(*Received 14 March 2015; accepted 16 May 2015*)

Crowdsourcing is a new approach which obtains information or input for a particular task by enlisting the services of the crowd. In recent crowdsourcing applications, the hybrid human–computer approach has been widely studied, to take advantage of both human beings and computers. In this paper, we propose a novel such application: blood typing for people in a family. We propose the *BloodTyping* method. It selects some members to take medical blood type tests, and to determine other family members' blood types, based on the inheritance rules. The aim is to reduce the number of medical tests, and thus, lower the cost. We extract rules for both $+$*induction* and $-$*induction*. The former is to predict children's blood types from parents', and the latter is to backward-induce a parent's blood type, given those of children and the other parent. Different combinations of blood types can induce different results: some may be an exact blood type, while others are composed of several blood types. Our method is optimised by conducting the cases which generate exact blood types first. The order is guided by extra-information via crowdsourcing, including the distribution of blood types with respect to the birthplace, and the personality, which may indicate some specific blood types. Taking a family with two parents and all children as a basic unit, the algorithm can be conducted simultaneously among different families in a decentralised way. The simulation results show that *BloodTyping* can significantly reduce the required number of blood tests.

Keywords: crowdsourcing; hybrid human–computer; blood typing; inducing blood type; inheritance rules

1. Introduction

Crowdsourcing is the practice of obtaining needed services, ideas or content by soliciting contributions from a large group of people (i.e. the crowd). In crowdsourcing applications, the hybrid human–computer approach has gained substantial interest [1,3,7,20,22], to take advantage of both human beings and computers. Human beings are generally good at identification and semantic analysis. They can be classified into two categories of ordinary crowds and the experts, according to whether the expertise is needed or not. Hiring experts usually takes a high cost. Computers can handle large-scale numerical analytics with a lower cost [10], but they are not efficient in performing some specific tasks. The cost of the crowd (crowdsourcing) falls in between experts and computers. Generally speaking, the more accurate the results we desire, the more cost we should pay. Properly incorporating experts, computers and the crowd can produce significant results economically, e.g. CrowdDB [7], CrowdSensing [21] and reCAPTCHA [19].

*Corresponding author. Email: jiangwenjun@hnu.edu.cn, wenjj8a@gmail.com

In this paper, we propose another novel application, i.e. non-emergent blood typing for all members in a family. Blood typing is the process of learning which blood type a person has, usually by taking a medical test. Knowing blood type can benefit people in multiple aspects, such as (1) keeping healthy (e.g. to eat or exercise as suggested); (2) preventing the risk for some specific diseases (e.g. [13] shows that blood type can actually predict the risk for heart disease) and (3) for urgent blood transfusion (http://en.wikipedia.org/wiki/Blood_transfusion) when medical blood type test resources are unaccessible. However, is it necessary to give each member a medical test, for the task of knowing all their blood types? Is there any way to reduce the number of medical tests, and to hasten the process? We resort to crowdsourcing for this end. Similar to the above-mentioned examples, in order to complete the task with higher efficiency and lower costs, it also needs a careful design of the incorporation of experts (i.e. medical test), computers and the crowd. We formulate this problem, and propose a hybrid blood typing framework which aims to minimise the medical tests for determining all the blood types.

We take a natural family (http://www.worldcongress.org/WCF/wcf_tnf.htm) that consists of two parents and their children as a basic unit, denoted as a *family unit*. Generally speaking, each person (adult) may be involved in two family units, i.e. where he/she takes the role of a child or a parent. Taking v_7 in Figure 1(b) for instance, he/she is a child in family unit 1, with v_1 and v_2 being the parents, and v_5 and v_6 being the siblings. Meanwhile, he/she is a parent in family unit 3, with v_8 being his/her spouse, and v_9 and v_{10} being the children. Additionally, there are several blood type groups. Here, we focus on the ABO blood groups, where there are four blood types: A, B, AB and O; and six gene types: AA, AO, BB, BO, AB and OO.

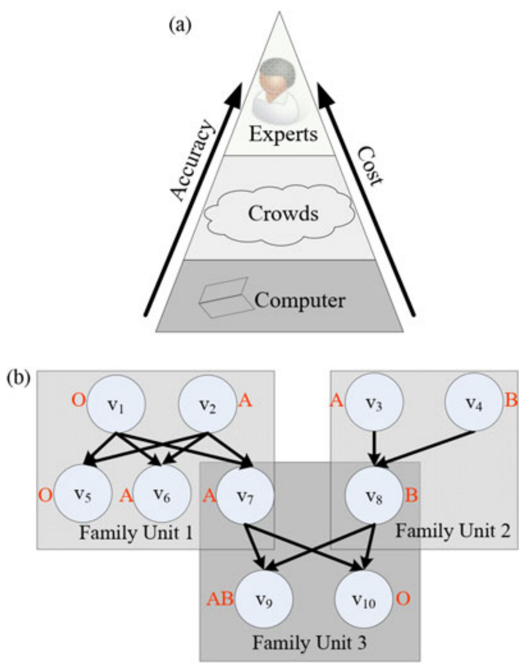

Figure 1. (a) The illustration of accuracy and cost in hybrid human–computer applications. (b) An example family tree with three family units. Each node is a person with his/her blood type at the side; each edge has a direction from parent to child.

Motivation example. Our goal is to reduce the number of medical blood type tests on members in a family (we will use test, blood test or medical test with the same meaning). We know that human blood types are inherited from parents. Using the inheritance rules, we can derive the possible blood types of offspring. However, usually there are several possible results for each child. Taking family unit 3 in Figure 1(b) for instance, the process of determining children's blood types is given in Figure 2(a). Suppose v_7 has type A blood, and v_8 has type B. The corresponding genotype for A is AA or AO, while genotype for B is BB or BO. Then, taking one gene from each parent, the child's blood type can be A, B, AB or O, each with an individual probability. Here, we cannot tell what the exact blood type will be. However, given the blood type of a parent and those of children, we can backward induce to get the other parent's blood type (Figure 2(b)). We can induce that v_8's blood type is B, given the others' blood types shown in the figure. Then, we can save one blood test for family unit 3.

Based on this, we can extract two key points for saving medical tests: (1) The *early detection* of forward induction or backward induction, that is, to save tests on children or on a parent. (2) The *blood typing order*, i.e. determining which members to test earlier in order to induce others' blood types.

Basic ideas and contributions. We propose a hybrid human–computer method, *BloodTyping*, to test or induce blood types for all members in a family. We extract induction rules for both predicting children's blood types from those of parents (i.e. $+$ *induction* representing forward prediction), and rules from children and a parent to

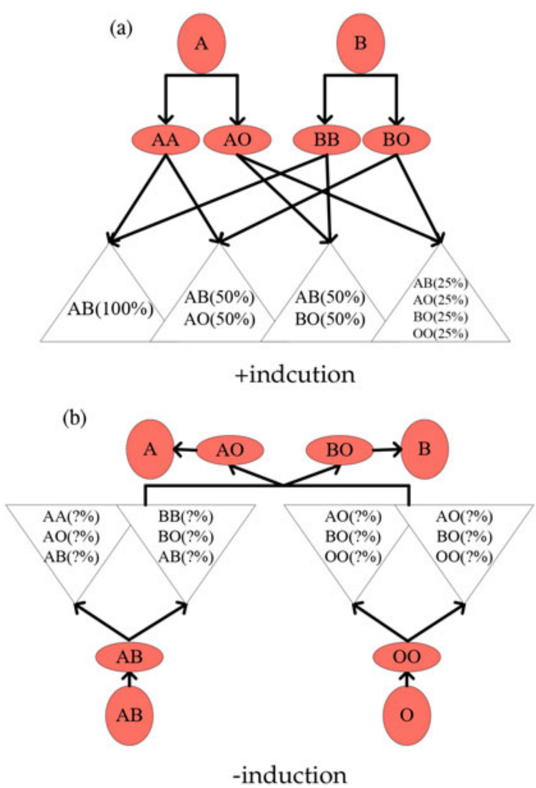

Figure 2. Inducing blood type for family unit 3: (a) forward (b) backward.

induce the other parent's blood type (i.e. $-$*induction* representing backward induction). We find that, if only considering blood types themselves (neglecting their distributions), the optimal blood typing order is to first test type O, then test AB, finally test A and B. In addition, for the children's blood type set, we care more about new blood types. Therefore, if we can estimate who are type Os, or which members may have different blood types, the saving chance will be increased. To this end, we incorporate crowdsourcing, to gain information such as the birthplaces, and the personalities which may indicate some specific blood types [6, 16]. This information can help estimate the possible blood types and can guide the blood typing order. Additionally, we differentiate three types of family relations – *intermarriage families*, *independent families* and *hierarchical families*. We also design a decentralised blood typing algorithm, which can be conducted simultaneously among different families.

Our contributions are manifold:

(1) As far as we know, we are the first to propose the novel application of hybrid human–computer blood typing for a family. Our work can be used for universal medical assistances, or blood type statistics, especially in some specific areas or scenarios where there are a limited number of medical resources.
(2) We extract rules for both $+$ induction and $-$ induction, to induce blood types and save the medical tests. During the extraction, we find that the earlier determination of types O and AB can benefit the savings of medical tests.
(3) We incorporate crowdsourcing to guide the blood typing order. To be specific, we first estimate the possible blood types by the birthplace and personality, then we select a minimal subset of members to take medical tests.
(4) We devise a decentralised blood typing algorithm to reduce the time cost.
(5) We conduct extensive simulations to validate the effects of the proposed method. It shows that the number of medical tests can be reduced significantly.

The remainder of this paper is organised as follows: we formulate our problem in Section 2 and then propose a blood typing framework in Section 3. Section 4 discusses the optimal blood typing order. Section 5 devises a decentralised blood typing algorithm. An experimental study is presented in Section 6. We cover related work in Section 7, and present our conclusion and future work in Section 8.

2. Problem formulation

In this section, we first formally describe the system settings. Then, we define two types of inductions of $+$ induction and $-$ induction (Section 2.1), and introduce how we incorporate crowdsourcing to help determine the order of blood typing (Section 2.2). Finally, we formulate the problem we address. Notations used in this paper are described in Table 1.

A BloodTyping system is constructed based on a family tree $G = (V, E)$, with $V = \{v_1, \ldots, v_n\}$ being the node set and E being the directional edge set. Each node is a member of the family. Each edge e_{ij} indicates that v_i is a parent of v_j. The blood type of a node v is denoted as t_v, while the blood type set of nodes in C is denoted as T_C. The aim is to induce some members' blood types based on the inheritance rules.

2.1 + Induction and − induction

We first give the definitions of $+$ induction and $-$ induction. Then, we extract the rules to guide the blood typing process.

Table 1. Notations.

Symbol	Description
$G=(V,E)$	A family tree a family unit with node set V And edge set E
F	A family unit
t_v	Blood type of node v
T_C	Blood type set of node set C
$P_v(t)$	Probability that v's blood type is t
k	Maximum number of children in a family
ω	$\{p_1, p_2, c_1, \ldots, c_n\}$, members with tentative order
E^+	Expected saving number of $+$ induction
E^-	Expected saving number of $-$ induction

Definition 1: $+$ *induction*. It is the process used to predict children's blood types, given those of parents.

Definition 2: $-$ *induction*. It is the process of predicting a parent's blood type, given those of children and the other parent.

It is worth noting that, given only children's blood type, we may induce the blood type combination of two parents. However, we cannot distinguish which one is the father's and which one is the mother's. Therefore, in $-$ induction, we need a parent's blood type as the known condition. As a result, we can save at most one medical test using $-$ induction.

For the $+$ induction, Table 2 shows the probabilities that a child has some blood types, given those of two parents. We can see that, there are a total of 10 possible combinations for parents' blood types. Among them, only the first case, O and O, can produce children with one exact blood type. This occurs in 1 out of 10 cases; the proportion is $1/10 = 10\%$.

Table 3 shows the results of $-$ induction. The first column depicts the blood type combinations of children. For the four ABO blood types, each may or may not occur in children, thus there are a total of $2^4 = 16$ cases. The first row depicts the four possible blood types of a parent. The remainder in the table is the induced possible blood type of the other parent. We can see that there are a total of $16 \cdot 4 = 64$ cases. Among them, 17 cases cannot happen (represented by " $-$ " in the table). And 15 out of 47 cases can be exactly induced. The proportion is $15/(64 - 17) \approx 31.915\%$.

The above comparison indicates that $-$ induction has more instances to save medical tests than $+$ induction. However, $+$ induction may actually save more tests than $-$ induction for two reasons: (1) $+$ induction can save many tests when there are many

Table 2. Blood types by $+$ induction.

Parents\child	A (%)	B (%)	AB (%)	O (%)
{O, O}	0	0	0	100
{A, O}	75	0	0	25
{A, A}	93.75	0	0	6.25
{A, B}	18.75	18.75	56.25	6.25
{A, AB}	50	12.5	37.5	0
{B, O}	0	75	0	25
{B, B}	0	93.75	0	6.25
{B, AB}	12.5	50	37.5	0
{AB, O}	50	50	0	0
{AB, AB}	25	25	50	0

Table 3. Blood types by $-$induction.

Children\parent	A	B	AB	O
{O}	A, B, O	A, B, O	–	A, B, O
{AB}	B, AB	A, AB	A, B, AB	–
{*AB*, O}	B	A	-	–
{B}	B, AB	A, B, AB, O	A, B, AB, O	B, AB
{*B*, O}	B	A, B, O	–	B
{B, AB}	B, AB	A, AB	A, B, AB	–
{*B*, *AB*, O}	B	A	–	–
{A}	A, B, AB, O	A, AB	A, B, AB, O	A, AB
{*A*, O}	A, B, O	A	–	A
{A, AB}	B, AB	A, AB	A, B, AB	–
{*A*, *AB*, O}	B	A	–	–
{*A*, *B*}	B, AB	A, AB	A, B, AB, O	AB
{*A*, *B*, O}	B	A	–	–
{A, B, AB}	B, AB	A, AB	A, B, AB	–
{*A*, *B*, *AB*, O}	B	A	–	–

children of type O parents. All of them will have type O blood, and will save medical tests. Meanwhile, $-$induction can save at most one test on a parent. (2) Within different population, there are varying blood type distributions. For instance, if type O represents 50% of the population, then the chance of an O, O combination will be quite large, indicating that $+$induction may occur frequently. Meanwhile, other blood types (or combinations) will take a small proportion, leading to the fewer occurrence of $-$induction.

Based on the inheritance rules on blood types (see Appendix or http://en.wikipedia.org/wiki/ABO_blood_group_system), we extract the following induction rules:

Rule 1. Only when both parents are with blood type O can the children's blood type be predicted exactly.
Rule 2. If the children's blood type set contains {O}, then, the parents cannot be type AB; and vice versa.
Rule 3. If the children's blood type set contains {AB, O}, then, the parents must be a type A (AO) and a type B(BO).
Rule 4. If the children's blood type set is $\{X, \mathrm{O}\}$ ($X =$ A or B), and parents are one X and another O, then, determining type O can help to induce the other, but not vice versa.
Rule 5. If a family has no child, no medical test can be saved.
Rule 6. If a family has only one child, use $+$induction only, since no medical test can be saved in $-$induction with a child.

Rule 1 can be applied to $+$induction. In this case, there is only one possible blood type for each child, that is O. The case is very useful for that, as long as we know the two parents' blood types, all children's blood types can be determined. Therefore, a type O parent deserves an earlier test. Rules 2, 3 and 4 can be applied to $-$induction. Only the cases in which children's blood types contain O do we have the chance to save medical tests (of a parent). Therefore, a type O child also deserves an earlier test. Rule 4 suggests the earlier test of a type O parent in $-$induction. Rules 5 and 6 describe two special cases of no-child and one-child families.

These rules, and the two tables, indicate that type O blood has a large impact on the final savings of medical tests. In addition, types A and B cannot be distinguished from each other. Therefore, intuitively, the optimal order is first to test type O, then type AB and finally A and B. However, the optimal order requires knowing the real blood types upfront, which cannot be achieved in reality. To solve this issue, we propose incorporating crowdsourcing approaches to estimate the possible blood types, so as to guide the blood typing order.

2.2 Crowdsourcing

We want to incorporate crowdsourcing to estimate the possible blood types. As far as we know, the most common study about blood types is the distribution in different areas, and the relation between blood type and personality. Due to the dynamic property of both blood type distribution and personality, there is no definite conclusion about that so far. However, those studies can at least give us some suggestions.

Many crowdsourcing platforms (e.g. Amazon Mechanical Turk (AMT)) provide Application Programming Interfaces (APIs) for calling workers to complete micro-tasks (human intelligent tasks [HITs]) [20]. To gain the information about a person's birthplace and personality, we can invite him/her to join the investigation, and answer simple questions such as (1) "In which area were you born?" (2) "Which description fits your personality the best?" Note that we will not predict a person's blood type directly by the answers. We only estimate his/her possible blood type by considering the birthplace or the personality. Based on the possible blood type, an initial blood typing order can be determined. Then, some members will take real medical tests, and the following order will be dynamically updated according to the real results.

2.3 Problem description

To process blood typing, we first use crowdsourcing to generate a candidate set of testing nodes. The goal of our work is to get all nodes' exact blood types. In our setting, some nodes will be selected by crowdsourcing results, and will take real medical tests; others' blood types will be induced using induction rules that we have previously extracted. We call the former 'crowdsourced' (tested) nodes, and the latter 'induced' nodes. Typically, it is more expensive to take a medical blood type test (about 10 dollars per person) than to perform a crowdsourcing task (about 25 cents per HIT). Thus, there is a financial incentive to minimise the number of medical tests. Formally, we define our problem as follows.

Definition 3: *Blood typing problem*. Given a set of family units where the blood types of all members need to be determined, our goal is to select a minimum number of nodes who will take real blood type tests, such that for the other nodes, their blood types can be induced.

3. Bloodtyping: the framework

In this section, we propose a BloodTyping framework. Our framework takes as input a set of family units for which the blood types of all members need to be known. People who are involved in this work will be asked to take simple crowdsourcing tasks. They will choose birthplaces from a series of given areas, and will answer some personality-related questions, so as to estimate their possible blood types. They are then sorted by these types. Some of them will be selected to take medical tests to determine their blood types, while others will be induced by $+$ induction or $-$ induction. In this paper,

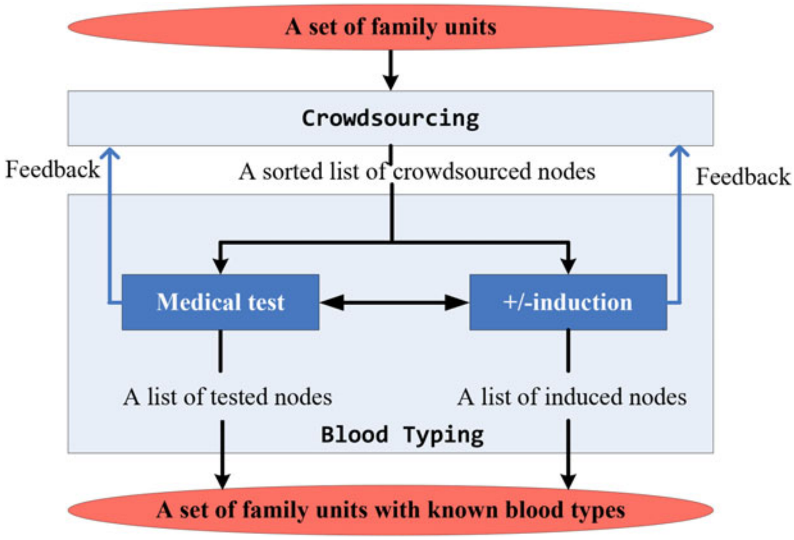

Figure 3. The hybrid framework.

we take both medical blood type tests and $+/-$ induction as blood typing. As shown in Figure 3, our framework mainly consists of two components, crowdsourcing and blood typing. Their details will be described in the following two subsections. We focus on studying the optimal blood typing order, and how to incorporate crowdsourcing to guide the order.

3.1 Crowdsourcing component

As mentioned before, we find that the blood typing order will affect the number of induced users and tested users. Based on this observation, in our framework, the crowdsourcing component attempts to identify the optimal blood typing order to minimise the number of tested nodes, and maximise that of the induced nodes. Thus, it takes as input a set of family units without knowing blood types, and outputs a sorted list of nodes, as shown in Algorithm 1. For the nodes who are parents in a family, we select possible type Os first. For those who are children, we prefer possible type Os and ABs; However, we select each type only once. For instance, we can first select a child with possible type O; for the remainders, each time, we will select a child who has a different possible blood type than that of previously selected children. In this way, we can determine the children's blood type set earlier, and thus induce a parent's blood type earlier. In Algorithm 1, lines 1–2 take a time complexity of $O(n)$, where n is the total number of nodes. Line 4 takes a constant time, because we only compare who is more possibly type O for two parents; lines 5–6 take $4 \cdot O(n)$, because we only select at most four representatives for children. Therefore, the total time complexity is $O(n)$.

A blood typing order can be taken as a sorted list of users, denoted by $\omega = \{p_1, p_2, c_1, \ldots, c_n\}$, where the first two are the parents, and the remainder ones are the children; p_2's order is not determined. Among all children, c_i's $(2 \leq i \leq n)$ blood type will be either tested or induced after c_{i-1}. Take family unit 3 in Figure 1(b) for instance, an optimal order is $\{v_7, v_8, v_{10}, v_9\}$. It means that the blood types of v_7, v_{10} and v_9 will be tested, and v_8's type will be induced (if possible), or finally tested (if it cannot be induced).

Algorithm. 1 Crowdsourcing (F)

Input: F, a family unit for blood typing.
Output: F', a list of members with tentative order.

for each node v in F **do**
Crowdsourcing for birthplace and personality. Using this information, estimate v's possible blood type.
Add nodes into F' based on their possible blood types:
Step 1: put a parent who is more like type O as the first, and another as the second.
Step 2: for children in a family, add the possible type O and type AB ones before others.
Step 3: add the child who has a different possible blood type than the previously selected children.
return $F' = \{p_1, p_2, c_1, \ldots, c_n\}$.
Note that the order is a tentative one and will be determined dynamically in Algorithm 2.

3.2 Blood typing component

In the blood typing component, we select some nodes to take medical tests and induce others' blood types by using $+/-$ induction. Without loss of generality, we assume that the medical test results are always correct. Then, the main task is to improve the efficiency of blood type induction.

Algorithm 2. BloodTyping (F')

Input: F', the resulting set of Algorithm 1.
Output: T/I, the set of tested/induced members.

//for a family having no child
if $F' = \{p_1, p_2\}$ **then**
  return $T = \{p_1, p_2\}$, $I = \emptyset$.
//for a family having one child
if $F' = \{p_1, p_2, c_1\}$ **then**
  Test parents p_1, p_2prime; blood types.
  if $t_{p_1} =$ O and $t_{p_2} =$ O **then**
   return $T = \{p_1, p_2\}$, $I = \{c_1\}$.
  else return $T = \{p_1, p_2, c_1\}$, $I = \emptyset$.
//for a family having more than one child
Test p_1's blood type.
if $t_{p_1} =$ O and p_2 is possibly type O **then**
  Test p_2's blood type.
   if $t_{p_2} =$ O **then**
   return $T = \{p_1, p_2\}$, $I = \{c_1, \ldots, c_n\}$.
   else return $T = \{p_1, p_2, c_1, \ldots, c_n\}$, $I = \emptyset$.
else //Begin to test children's blood types.
   for each child c in F' **do**
  Test c's blood type. $T \leftarrow c$.
  if $I = \emptyset$ **then**
  Induce p_2's blood type.
  if can induce an exact blood type **then**
   return $T = \{p_1, c_1, \ldots, c_n\}$, $I = \{p_2\}$.
   return $T = \{p_1, p_2, c_1, \ldots, c_n\}$, $I = \emptyset$.

Based on the sorted list resulting from the crowdsourcing process, we select some nodes to take medical tests. As shown in Algorithm 2, there are three cases according to the number of children. (1) A family with no children cannot save tests (lines 1–3). (2) A family having only one child may save a test only when both parents are type O, so we first test the parents' blood types (lines 4–9). (3) For a family with more than one child (lines 10–24), at least a parent p_1 is selected for testing (line 11). If his/her blood type is O, and the other parent p_2 is possible type O, then we test p_2's blood type (lines 12–13). If we get another O, then, all the children's blood types are O by the + induction rule (lines 14–15); otherwise, we cannot save medical tests (line 16). If the first parent is not O, or p_2 is not possibly type O, we will test the children and try to induce p_2's blood type (lines 17–24): each time we test a child's blood type in the order proposed in F' (line 19). Then we try − induction to see if p_2's blood type can be induced as early as possible (lines 20–23). The worst-case scenario is after all the children's blood types are tested, o_2's blood type cannot be induced (line 24). The total time complexity is O(n).

4. Optimal blood typing order

As we have mentioned before, different blood types take different roles in inducing others' blood types. In this section, we first provide two theorems to formally describe the optimal blood typing order. Then, we study how the estimated blood types will benefit the results, and how to incorporate crowdsourcing to determine the order of blood typing.

4.1 Two theorems

We propose two theorems, in which we only consider the blood types themselves and neglect their distributions.

Theorem 1: Considering only the blood types, the optimal blood typing order is to test type O first, and then AB, A/B.

Proof. We consider four cases of testing four blood types. In each case, we differentiate two subcases of whether we are testing a parent or a child. We can prove the theorem by calculating (1) the number of possible blood types that those cases may generate, denoted as α_i^y, where $i \in [1, 4]$, and y can be p or c, represents the testing of a parent or a child; and (2) the number of possible blood types they generate through combination with other types, denoted as β_i^y. The fewer possible blood types there are, the better the case is.

Case 1: test type O. (1) If we test a parent who has type O blood, then, the possible blood types for his/her children could be O, A and B. It cannot be AB. Therefore, $\alpha_1^p = 3$. When combining with other blood types, we can get $\beta_1^p = 2$ according to Table 2. (2) We can get the same results for parents if we test a child who has type O blood. Therefore, $\alpha_1^c = 3$. When combining with other blood types, we can get β_1^c, which can be any value in $\{1, 3\}$, according to Table 3.

Case 2: test type AB. (1) If we test a parent who has type AB blood, then, the possible blood types for his/her children could be AB, A and B. It cannot be O. Therefore, $\alpha_2^p = 3$. When combining with other blood types, we can get β_2^p, which can be any value in $\{2, 3\}$ (Table 2). (2) If we test a child who has type AB blood, we can get the same results for parents. Therefore, $\alpha_2^c = 3$. When combining with other blood types, we can get β_2^c, which can be any value in $\{1, 2, 3\}$ (Table 3).

Case 3: test type A. (1) If we test a parent who has type A blood, then, the possible blood types for his/her children could be O, A, B and AB. Therefore, $\alpha_3^p = 4$. When combining with other blood types, we can get β_3^p,which can be any value in {2, 3, 4} (Table 2). (2) If we test a child who has type A blood, the parents could be any ABO blood types. Therefore, $\alpha_3^c = 4$. When combining with other blood types, we can get β_3^c, which can be any value in {1, 2, 3, 4} (Table 3).

Case 4: test type B. Do the same process with case 3. Then, $\alpha_4^p = 4$, $\alpha_4^c = 4$, β_4^p is in {2, 3, 4}, and β_4^c in {1, 2, 3, 4}. By comparing α_i^y and β_i^y, we complete the proof. □

Theorem 2. Considering the role of a member in a family, for only the parents, Theorem 1 still applies. For parents and children, or only the children, an additional principle is to test different blood types first.

Proof. We consider the following two cases to show why testing different blood types first will be beneficial.

Case 1: for parents and children. As shown in Table 3 for − induction, for all the cases that a parent p's blood type can be exactly induced, p provides a genotype that some children have, but the given parent does not. It indicates that selecting to test a parent who may have a different genotype than his/her children will have a high probability of exact induction.

Case 2: for only children. Again, we consider − induction in Table 3. We can see that a single blood type of children cannot help induce parent's exact blood type. The more different blood types there are in children's combination set, the larger the probability that we can induce. □

4.2 Estimated blood types and expected savings

We discuss how the estimated blood type can impact the expected saving number of medical tests. Let p_1 represent a parent in a family, and p_2 be another; and $C = \{c_1, \ldots, c_k\}$ be the children. Suppose we can estimate parents' blood types as being O with probability $P_{p_1}(\mathrm{O})$ and $P_{p_2}(\mathrm{O})$. Then, the expected saving number of + induction will be as follows:

$$E^+ = P_{p_1}(\mathrm{O}) \cdot P_{p_2}(\mathrm{O}) \cdot k, \tag{1}$$

− induction is more complex because of the varying number of children. Taking $k = 2$ for instance, children's blood type set T_C can be {O}, {A}, {B}, {AB}, {AB, O}, {A, O}, {B, O}, {A, B}, {A, AB}, {B, AB}. Among them, only {AB, O}, {A, O}, {B, O}, {A, B} have the chance to induce an exact blood type. Together with a parent's blood type, a total of seven cases can save one medical test (Table 3): $\{AB, O\} \cup \{A\}$, $\{AB, O\} \cup \{B\}$, $\{A, O\} \cup \{B\}$, $\{A, O\} \cup \{O\}$, $\{B, O\} \cup \{A\}$, $\{B, O\} \cup \{O\}$ and $\{A, B\} \cup \{O\}$. Let j be the index of the seven cases ($j \in [1, 7]$). Taking the first case $\{AB, O\} \cup \{AA\}$ for instance, the expected saving number E_1^- can be calculated as $E_1^- = [P_{c_1}(\mathrm{AB}) \cdot P_{c_2}(\mathrm{O}) + P_{c_2}(\mathrm{AB}) \cdot P_{c_1}(\mathrm{O})] \cdot [P_{p_1}(\mathrm{A}) + P_{p_2}(\mathrm{A})]$. The total expected savings can be summarised as follows:

$$E^- = \sum_{j \in [1,7]} E_j^- \tag{2}$$

4.3 Crowdsourcing for guiding blood typing order

In the above subsection, we calculate the expected saving numbers, given the probability of some member being some blood type. Now, we will calculate this probability according to the birthplace and personality. It is worth noting that, heuristically, birthplace takes more of an effect on the parents than on the children, while personality takes more of an effect on the children than on the parents. The intuitive reason is that the two parents may come from two totally different places. While for children in the same family (although they may be born in different places), the birthplace should not impact them much. In addition, the personality of an adult is formed by many factors, rather than only the blood type, while that of a child can be taken as formed by birth (i.e. it has not been shaped by other factors).

Some prior information will be used by the Bayes equation [11, 17]. Again, consider a family with parents p_1 and p_2. Suppose in the area of p_1's birthplace, the probability of a person being type O blood is $P^{\mathrm{B}}_{p_1}(\mathrm{O})$. Similarly, we have $P^{\mathrm{B}}_{p_2}(\mathrm{O})$ considering p_2's birthplace, and $P^{p}_{p_1}(\mathrm{O})$ and $P^{p}_{p_2}(\mathrm{O})$ considering p_1's and p_2's personality, respectively. Now, we can calculate $P_{p_1}(\mathrm{O})$ and $P_{p_2}(\mathrm{O})$, the probability that p_1 is of type O and p_2 is of type O, respectively. Take p_1 for instance. Let e_1, e_2, e_3 be the events which are defined as follows. e_1: a person's birthplace belongs to area $\mathcal{A}$. e_2: a person's personality belongs to category $\mathcal{B}$. e_3: a person's blood type is O. Let $P(e_3|e_1)$ represent the probability that a person born in area $\mathcal{A}$ has blood type O and $P(e_3|e_2)$ be the probability that a person whose personality falls into category $\mathcal{B}$ has blood type O. In addition, let $P(e_3)$ be the probability that any person in the world is type O (representing the percentage of all population in the world have type O blood). Suppose we know $P(e_3|e_1) = 0.6$, $P(e_3|e_1) = 0.4$, $P(e_3) = 0.45$. We also assume that e_1 and e_2 are independent events, then, $P(e_1e_2) = P(e_1)P(e_2)$. For a person f who is born in area $\mathcal{A}$, and has a personality in category $\mathcal{B}$, the probability that p_1's blood type is O, $P_{p_1}(\mathrm{O})$ is calculated as follows:

$$
\begin{aligned}
P(e_3|e_1e_2) &= \frac{P(e_1e_2e_3)}{P(e_1e_2)} = \frac{P(e_1e_2|e_3)\cdot P(e_3)}{P(e_1)\cdot P(e_2)} \\
&= \frac{P(e_1|e_3)\cdot P(e_2|e_3)\cdot P(e_3)}{P(e_1)\cdot P(e_2)} \\
&= \frac{\frac{P(e_3|e_1)\cdot P(e_1)}{P(e_3)}\cdot\frac{P(e_3|e_2)\cdot P(e_2)}{P(e_3)}\cdot P(e_3)}{P(e_1)\cdot P(e_2)} \\
&= \frac{P(e_3|e_1)\cdot P(e_3|e_2)}{P(e_3)} \\
&\approx 0.5333
\end{aligned}
\tag{3}
$$

Similarly, we can calculate $P_{p_2}(\mathrm{O})$, $P_{c_i}(X)$ $(X = \{\mathrm{O}, \mathrm{AB}, \mathrm{A}, \mathrm{B}\})$. Moreover, we can further save some probability calculations for children. As mentioned before, we care more about the different blood types of children. Hence, we can take only part of them, i.e. representatives who have different personalities with each other, so as to get all possible blood types in children. After estimating each member's possible blood type, we can sort them by using Algorithm 1.

5. Decentralised bloodtyping

In this section, we first distinguish three types of relationships between two family units: intermarriage families, independent families and hierarchical families. Then, we present a decentralised bloodtyping algorithm that can be conducted simultaneously in different family units.

5.1 Relations between two family units

Definition 4: Intermarriage families. Two family units are called intermarriage if they have no common nodes, but a node in a unit is married to a node in the other unit.

Definition 5: Independent families. Two family units are independent if they have no common nodes and they are not intermarriage families. That is, their members are not relatives.

Definition 6: Hierarchical families. Two family units are hierarchical if they share a common node who is a child in one family and a parent in another family. The former is called the *parent family*, and the latter the *child family*.

Taking Figure 1(b) for instance, family units 1 and 2 are intermarriage families; family units 1 and 3, and 2 and 3 are hierarchical families, respectively.

In addition, if taking the set of three families in Figure 1(b) as a bigger unit, there may also be overlap between the units. In this case, we can adopt a dynamic locking scheme to ensure each calculation an atomic action while still supporting concurrency. The *two-phase locking* (2PL) [2] is such a scheme where consistency is guaranteed if it is *well formed* and *two-phase*. A computation is well formed if it (1) locks an object (e.g. a person) before accessing it, (2) does not lock an object that is already locked and (3) before it completes, unlocks each object it has locked. A computation is two-phase if no object is unlocked before all needed objects are locked.

5.2 Decentralised bloodtyping

In the decentralised BloodTyping algorithm, independent families can conduct blood typing (Algorithm 2) simultaneously. However, intermarriage families and hierarchical families should be treated carefully. Taking Figure 1(b) for instance, suppose we treat family units 1 and 2 first. For family unit 1, the best case is to test v_1 (who is possible type O), and subsequentially test all children, and v_2 can be induced. Family unit 2 has only one child, so parents are first tested, then the child is also tested since his/her blood type cannot be induced. Then, only one test (for v_2) can be saved. However, if we treat family units 1 and 3 first, we can save two tests for v_8 and v_2. Therefore, we can improve the savings by properly treating intermarriage and hierarchical families. We design the following additional rules for the decentralised algorithm.

Algorithm 3. Decentralised BloodTyping (F_1; F_2; F_3)

Input: F_1 and F_2 are intermarriage families; F_1 and F_3, F_2, and F_3 are hierarchical families.

Output: T, a set of tested nodes; I, a set of induced nodes.

1: Call Algorithm 1 for F_1, F_2 and F_3 simultaneously.

2: Apply *Rule 7* to F_1, F_2 and F_3, using the following steps (add tested nodes into T and induced nodes into I):

3: Step 1: select families with two possible type O parents. Test the parent who is most likely type O. If an O is achieved, test the other parent.

4: Step 2: Call Algorithm 2 (bloodtyping) to treat the new selected family units, using *Rule 8* and *Rule 9*.

5: **return** T, I.

Rule 7. Treat the family with two possible type O parents first.

Rule 8. For simultaneous treating of two hierarchical families, treat the child family first.

Rule 9. For simultaneous treating of two intermarriage families, treat the family with more children first.

Rules 7–9 are applied in Algorithm 3, to further save the medical tests. Rule 7 is used to give type O parents the priority, with the expectation of saving tests on all children in the family. Rule 8 gives higher priority to the child's family, with the expectation of saving a test on one parent, which may further save another test on a grandparent. Rule 9 gives the priority to the family with more children, which has a larger chance of saving tests.

5.3 Optimisation techniques

We propose two optimisation techniques to further enhance our decentralised algorithm, as follows.

Instant decision: recall our decentralised bloodtyping algorithm. It will first crowdsource all members of multiple family units to the crowdsourcing platform. After we have estimated the possible blood types of all the crowdsourced nodes, independent families can conduct the BloodTyping algorithm (Algorithm 2) simultaneously. Intermarriage and hierarchical families will be treated using *Rules 7–9* (Algorithm 3). Some family members will wait until others are tested or induced. Notice that we do not need to wait to decide the next-round tested or induced pairs. Instead, when some of the crowdsourced nodes are being tested or induced, we can utilise them instantly to treat the remainders.

Type O firs: As we have mentioned, in optimal blood typing order, type Os are suggested to be tested first. Therefore, for any members in intermarriage or hierarchical families, possible type Os can be tested first without waiting for others.

6. Experiment

In this section, we evaluate our BloodTyping method. The goals of simulation are to (1) examine and compare the effectiveness of + induction and − induction (i.e. the number of savings on medical tests), (2) examine the effectiveness of optimal blood typing order and (3) validate the advantage of incorporating crowdsourcing. Correspondingly, we divide the experiments into three parts as follows.

6.1 Expected saving numbers

We study how the (1) blood type distribution and (2) +/− induction can impact the expected saving numbers.

Data generating. Without loss of generality, we construct 10,000 virtual family units. (1) We first generate 10,000 males and 10,000 females; their blood types are set randomly according to BloodBook (www.bloodbook.com), which has collected the distribution of four ABO blood types in 50 different groups of people (www.bloodbook.com/world-abo.html#Maoris). (2) We randomly match a male and a female, to construct 10,000 pairs of parents. (3) We randomly generate children for each pair of parents. The blood type of each child is randomly determined by inheritance rules in Table 2. The number of children is set randomly in the range of [1] (http://www.worldcongress.org/WCF/wcf_tnf.htm).

Table 4. Blood type distribution in different data-sets.

	Data-set	A (%)	B (%)	AB (%)	O (%)
Real	D1	42	8	3	47
	D2	27	32	13	28
	D3	33	32	18	17
	D4	39	0	0	61
Synthetic	D5	25	25	25	25
	D6	10	10	30	50
	D7	30	30	30	10
	D8	15	15	50	20

As shown in Table 4, we generate eight data-sets. The blood distributions of D1–D4 are set according to four real people groups in BloodBook, while those of D5–D8 are set artificially.

The effects of + induction and − induction. Figure 4 shows the medical test saving numbers using + induction and − induction. We have three main findings: (1) In both real (D1–D4) and artificial (D5–D8) blood type distributions, the saving number of using + induction is approximately linear to the number of children. (2) The saving number of using − induction is increased with the number of children, and shows a diminishing return. That is, it first increases quickly, and reaches the highest speed at $k = 3$ or $k = 4$, then its increase slows down. (3) The saving number of + induction is far more than that of − induction, especially when it has a larger percentage of the type Os. This finding

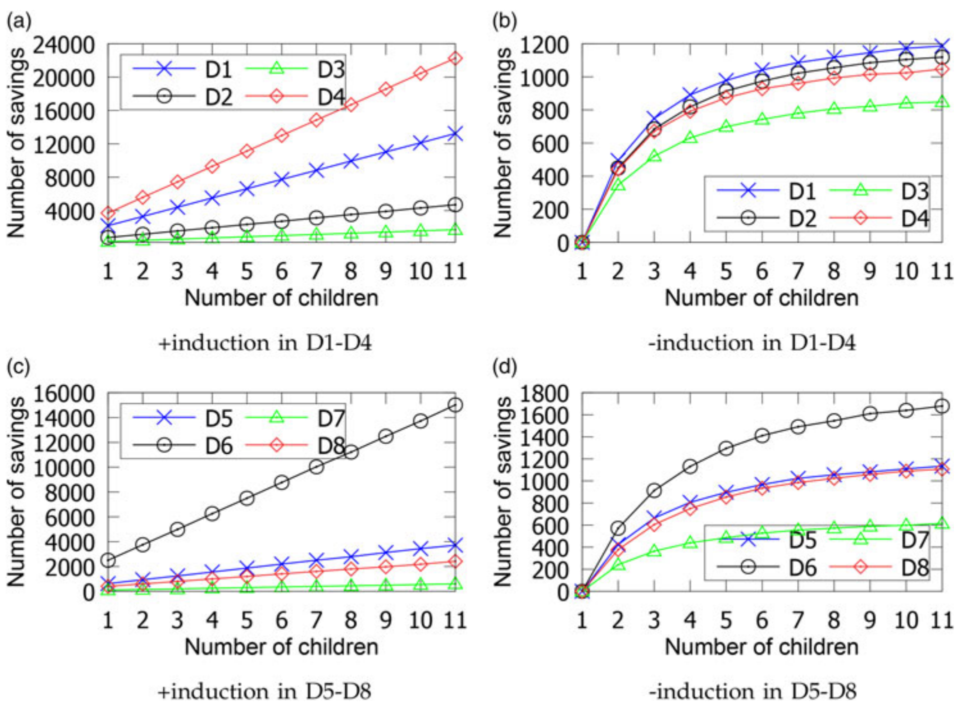

Figure 4. Number of medial test savings using + induction and − induction. (a) + induction in D1-D4. (b) − induction in D1-D4. (c) + induction in D5-D8. (d) − induction in D5-D8.

validates our analysis in +/− induction. That is, although there is only one case in which + induction can save, it can save more because the case has a high likelihood of occurring, and once it occurs, the saving number is larger (if there are many children).

The effects of blood type distribution. From Figure 4, we can also gain some findings with respect to the different effects of different blood type distributions. (1) Type O increases savings in + induction: in Figure 4(a), D4 makes the largest savings, then D1, D2 and D3, which is consistent with the order of type O's percentage. That is, the more type O people there are, the more savings + induction can make. Figure 4(c) also validates this point. (2) The combination of $\{X, \mathrm{O}\}$ (X can be A, B or AB) helps savings in − induction: in Figure 4(b), D1 makes the largest savings, then D4, D2 and D3. In Figure 4 (d), the order is D5, D6, D7 and D8. The order is different from that in + induction. It indicates that for − induction, the balance of the combinations of $\{X, \mathrm{O}\}$ (X can be A, B or AB) takes more effects than only a single type O. Recall what we have mentioned in − induction, the combinations can lead to exact induction. The percentage distribution in Table 4 and the savings in Figure 4(b),(d) validate this point.

6.2 Blood typing with order

In this subsection, we will conduct blood type inductions using the data-sets in Table 4. The goal is to gain the real medical test saving numbers. As we have proven before, the optimal blood typing order is O, AB, then A and B. Since we do not know the blood type when conducting the BloodTyping algorithm, we assume that each member has a larger probability of being some blood type than other blood types, and the probability is already known. This information will be used to guide the blood typing process.

Evaluation metrics. Similar to [15], we use two metrics, i.e. the *coverage* and the *accuracy*. Coverage is the ratio of real induction compared to the total cases that can be induced. Accuracy represents the ability of inducing a user's blood type or not, as in the following: Precision= $\sum \mathrm{A}_X \cap \mathrm{B}_X / \sum \mathrm{B}_X$, Recall= $\sum \mathrm{A}_X \cap \mathrm{B}_X / N$, Fscore=2 · Recall · Precision/(Recall + Precision), where A_X is the number of people whose real blood type is X (($X \in \{\mathrm{A}, \mathrm{B}, \mathrm{AB}, \mathrm{O}\}$)), and B_X is the number of people whose blood type is induced as X, N is the total expected savings that can be made. The Fscore metric is used to measure the accuracy using Recall and Precision jointly.

Methods of comparison. We compare two methods: (1) Random method, which randomly chooses one parent to take a blood type test and randomly conducts + induction or − induction. (2) Greedy method, which uses the most probable blood type as the estimated type, to guide the blood typing order. We assume for at most a percentage of p users, that their possible blood types are the same as the real blood type. Other users' possible blood types are randomly set using the remaining blood types other than the real type.

The effects of different methods. Figure 5 shows the comparison of different methods (random and greedy with $p = 0.5$) in two real data-sets of D1 and D2, and two synthetic data-sets of D5 and D6. Figure 6 shows the results of the greedy method with $p \in [0.5, 0.9]$. The results indicate that the greedy method beats the random method in both coverage and accuracy, even when at most half of the blood types are right, i.e. $p = 0.5$. (1) For the coverage, the random method takes a value that is close to (but less than) 0.5, while the greedy method with $p = 0.5$ has a coverage that is more than 0.5 and is close to 0.6. The improvements range from 1.6% to 34.56%. (2) For the accuracy, the greedy method makes an improvement over the random method, from 1.8% to 36.96%. (3) The improvements in D2 and D5 are larger than those in D1 and D4. We analyse that the

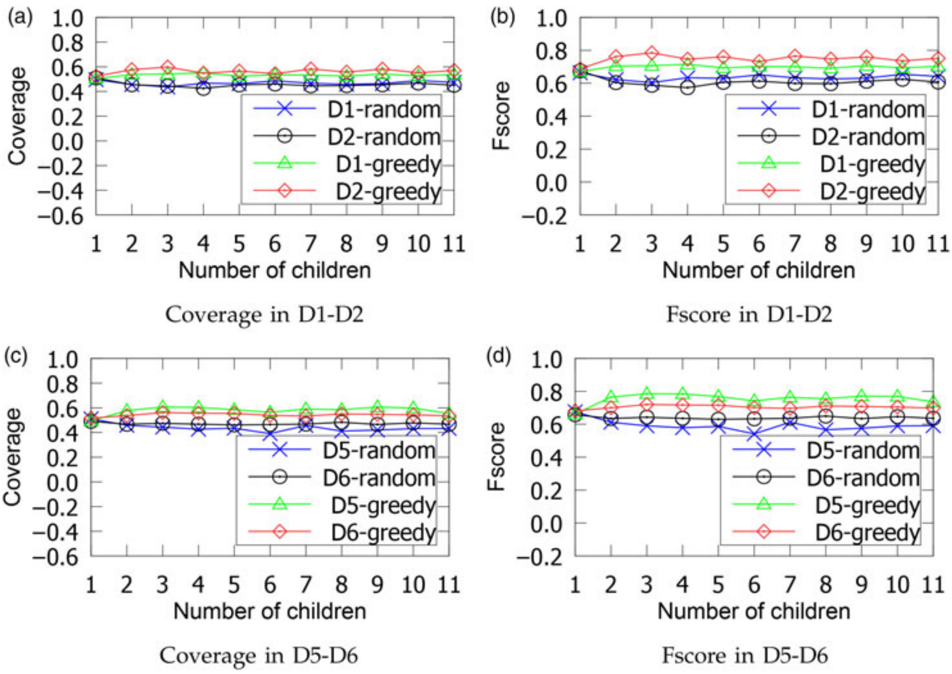

Figure 5. Comparison of coverage and accuracy (Fscore) with two methods (random and greedy): (a) and (b) in D1 and D2, (c) and (d) in D5 and D6.

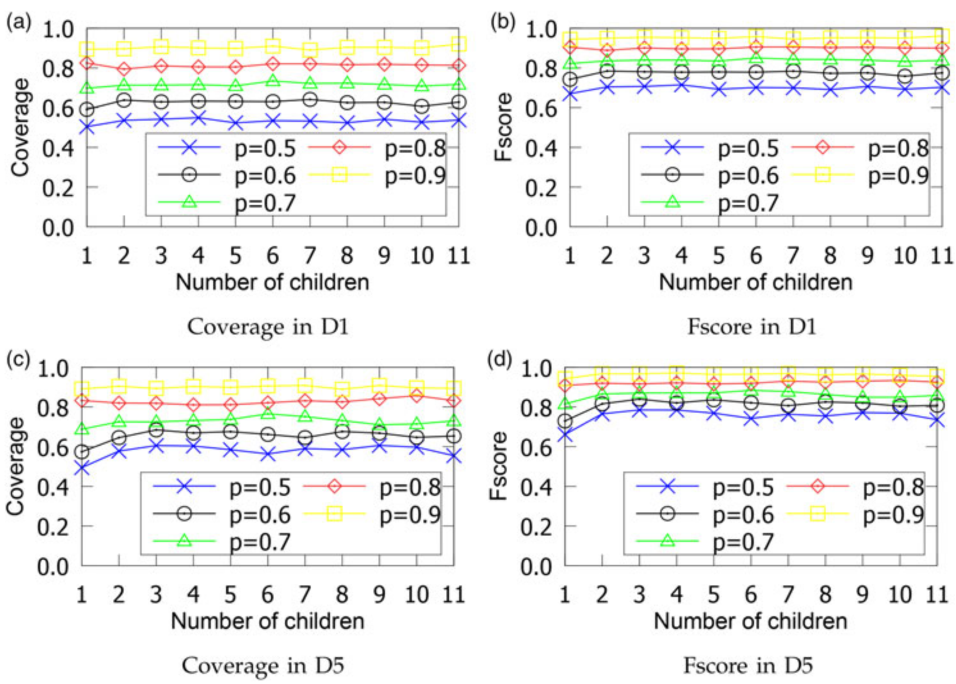

Figure 6. Comparison of coverage and accuracy (Fscore) using the greedy method with different percentages of correct estimation (p). (a) Coverage in D1. (b) Fscore in D1. (c) Coverage in D5. (d) Fscore in D5.

reason lies in the balance of different blood types, which are better in the former two. (4) The improvements hit the lowest point when the number of children k is 1. It increases significantly when k is around 3–6. After that it remains stable. (5) Figure 6 show that the coverage and accuracy change almost proportionally when p changes from 0.5 to 0.9, in all data-sets. It indicates that, if we can estimate blood type with a higher probability, the performance will be significantly improved.

Figure 7(a),(b) shows the saving ratios, i.e. real/expected saving numbers. We can see that for types AB and O, the ratios are 1 with both methods. Meanwhile, for types A and B, the random method fails in some cases. Those findings show the effectiveness of the greedy method. Figure 7(c),(d) shows the percentage of savings with respect to different blood types in data-set D1. We have several findings: (1) one-child families cannot save any blood types other than type O. (2) Type O is the most prevalent in both random and greedy (with $p = 0.5$) methods. It indicates that + induction makes more savings. (3) Type A takes a larger proportion in the greedy method than in the random method. (4) With the increase of children, type O's percentage increases. It also indicates the effects of + induction. That is, two type O parents can lead to more savings if they have more children. Those findings validate the reasonability of our proposed optimal blood typing order.

6.3 Incorporating crowdsourcing

We conduct two groups of experiments. One is for the area distribution, which is suitable for distinguishing parents' possible blood types. The other is for the personality, which is

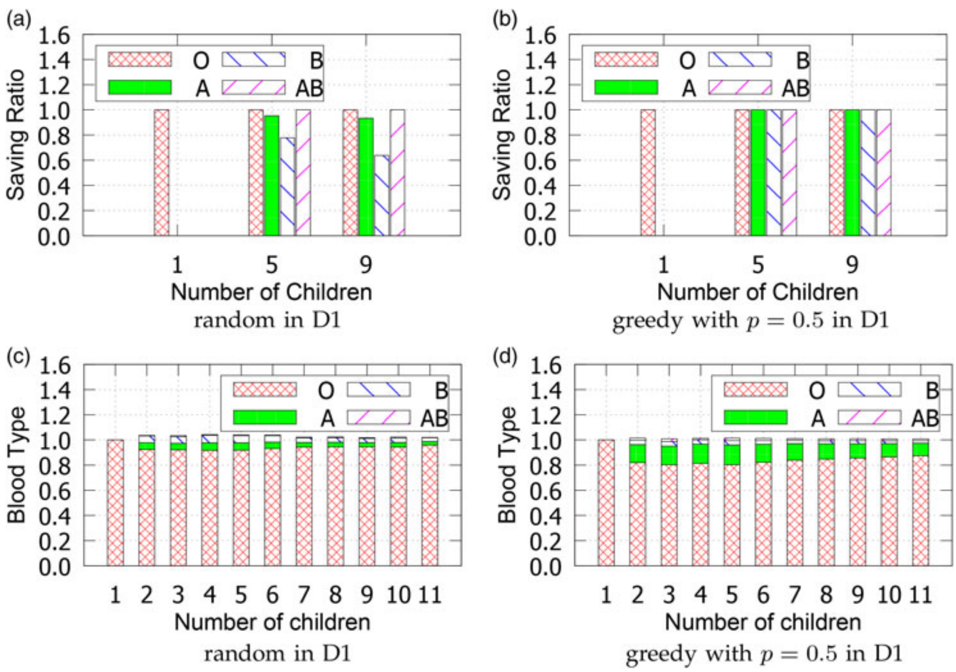

Figure 7. Saving ratios (real/expected) of four blood types in D1, and percentages of four blood types in real saving: (a), (c) random method, and (b), (d) greedy method.

Table 5. Typical personality with respect to blood type.

Type	Personality
A	Artistic, shy, conscientious, trustworthy, sensitive
B	Goal-oriented, strong-minded
AB	Split personalities, outgoing and shy, confident and timid
O	Outgoing, very social

suitable for distinguishing those of children. Table 5 shows the typical personalities corresponding to different blood types.

Data settings. Without loss of generality, we assume that there are four areas where the blood type distributions are the same with the data-sets D1–D3 and D8. That is to say, for people who are born in the area of D1, their blood types are O for 47%, A for 42%, and so on. Therefore, for a person born in one of the four areas, the possible blood types are O, B, A, AB, respectively. We consider two cases in which the population (i.e. the parents) is distributed (1) uniformly (i.e. 1/4) and (2) non-uniformly (e.g. 0.2, 0.3, 0.4, 0.1) in each area. Our method can be easily extended when applied to practical cases, in which there may be more areas to be considered. Similarly, we consider four types of personalities for children corresponding to four blood types, either uniformly or non-uniformly distributed.

The effects of the area distribution. We compare four settings of population distribution, in which the percentages for D1–D3 and D8 are set as follows: uniform = (0.25, 0.25, 0.25, 0.25), non1= (0.2, 0.3, 0.4, 0.1), non2= (0.1, 0.2, 0.3, 0.4) and non3 = (0.4, 0.3, 0.2, 0.1). Figure 8 shows the resulting coverage and accuracy. We can see that non3 gets the best performance, followed by uniform, non2 and non1. We analyse the reason, and again we find that the performance is related to the blood type distribution in each area. A higher percentage of O or balance of combination $\{X, \mathrm{O}\}$

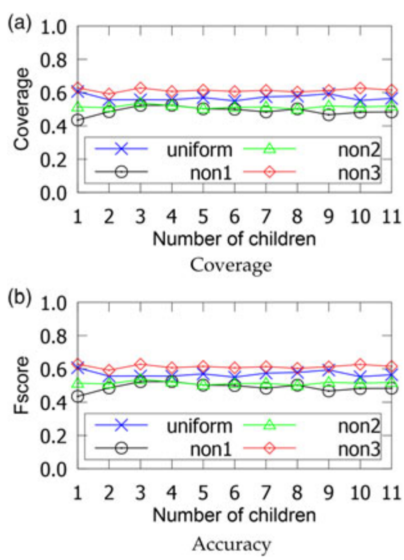

Figure 8. The effects of incorporating area distribution from crowdsourcing. (a) Coverage (b) Accuracy.

$(X = \mathrm{A}, \mathrm{B}, \mathrm{AB})$ leads to a better performance. We get similar results when testing the effects of personality. Due to space limitation, we do not display the results here.

6.4 Summary of experiments

We conduct many simulations to validate the effects of the proposed BloodTyping method. Multiple factors including the blood type distribution, the blood typing order (random or greedy) and the crowdsourcing are tested. The main findings are (1) $+/-$ induction can significantly save the medical blood type tests. In addition, $+$ induction saves more than $-$ induction. (2) Comparing the greedy and the random methods, the estimated blood types do help improve the performance, indicating the necessity of incorporating crowdsourcing for estimation. (3) The percentage of O impacts the savings of using $+$ induction, while the balance of the $\{X, \mathrm{O}\}$ (X can be A, B, or AB) combination impacts that of $-$ induction.

7. Related work

We first briefly review related work in crowdsourcing, then we describe the existing work on blood type distribution, and the relation between blood type and personality.

Hybrid human–computer and Crowdsourcing. Crowdsourcing has been used in many tasks such as crowdsensing [21], [9], [5] language translation [4], [14] and database operations [20], [12], [8]. Among them, [20] proposed an interesting work, to leverage the transitive relations for crowdsourced joins. For instance, if o_1 matches with o_2, and o_2 matches with o_3, then they can deduce that o_1 matches with o_3 without needing to crowdsource (o_1, o_3). They adopted a hybrid transitive relations and crowdsourcing labelling framework: (1) use machines to generate a candidate set of matching pairs, (2) ask humans to label the pairs in the candidate set as either matching or non-matching and then (3) use transitive relations with machines to deduce other pairs. To crowdsource the minimum number of pairs, they proof the optimal labelling order and devise a parallel labelling algorithm to efficiently crowdsource the pairs following the order. Wang's work gave us insights into inducing blood types, which is more difficult and challenging. In our work, we cannot ask the crowd to report their blood types, since in that case, they will have to take medical tests. Alternatively, we can learn the birthplace and the personality of the crowd by asking simple questions. The results can be used to decide (1) which ones are likely to have some specific blood type, and (2) which members have different personalities, and thus, may have different blood types.

Background: ABO blood group. Everyone has an ABO blood type, i.e. A, B, AB or O. The four types are called phenotypes, because they can be explicitly tested. In fact, There are six genotypes in the ABO blood group, i.e. *AA*., AO, BB, BO, AB, and OO. Each person has a specific blood genotype from the six. Blood types are inherited from parents. Each biological parent donates one of two ABO genes to his/her child. The A and B genes are dominant, and the O gene is recessive. For example, if an O gene is paired with an A gene, the blood type will be A, i.e. genotype AO shows type A in the medical test. Similarly, BO shows type B. Table 6 depicts the rules of possible genotypes of offspring, given the parents' genotypes. From the table, we can derive the probability of a child having a specific blood type.

Blood type. The distribution patterns of blood types are complex. Taking type A for instance (Figure 9), about 21% of all people share the A allele. The highest frequencies of A are found in small, unrelated populations, especially the Blackfoot Indians of Montana

Table 6. The rule of blood genotype by inheritance.

Parent genes		A		B		AB	O
		AA	AO	BB	BO	AB	OO
A	AA	AA	AA, AO	AB	AB, AO	AA, AB	AO
	AO	AA, AO	AA, AO, OO	AB, BO	AO, BO, AB, OO	AA, AB, AO, BO	AO, BO,
B	BB	AB	AB, BO	BB	BB, BO	AB, BB	BO
	BO	AO, AB	AB, AO BO, OO	BB, BO	BB, BO, OO	AB, AO, BB, BO	BO, OO
AB	AB	AA, AB	AA, AO, BO, AB	AB, BB	BB, BO, AB, AO	AA, BB, AB	AO, BO
O	OO	AO	AO, OO	BO	BO, OO	AO, BO	OO

(30–35%), the Australian Aborigines (many groups are 40–53%) and the Lapps, or Saami people, of northern Scandinavia (50–90%). *Relation with personality*. There is a long history of the study on blood type and personality (http://en.wikipedia.org/wiki/Blood_types_in_Japanese_ culture) [18]. Some key personality features of people with different blood types are summarised in Table 5. However, both the distribution and personality are not strictly defined. Therefore, they can only be used for the guidance of the blood typing order. Additionally, a website has developed a blood type calculator (www.endmemo.com/medical/bloodtype.php). However, in the backward calculation, they only consider one child together with a parent to calculate the other parent's possible blood type. In our work, we consider all possible combinations of children's blood types, and we focus on predicting the exact type.

8. Conclusion and future work

In this paper, we propose a novel hybrid human–computer application-blood typing for members in a family. We present the BloodTyping method, to test or induce the ABO blood types, based on the inheritance rules. The goal is to reduce the number of medical tests, and thus reduce the financial cost. We extract induction rules for both + induction and − induction; + induction is used to predict children's blood types from parents, and − induction is used to determine a parent's blood type, given the blood types of the children and the other parent. We incorporate crowdsourcing to estimate the possible blood types. According to our pre-defined selection rules, some members are selected to take medical blood type tests, while others' will be induced. In addition, we differentiate

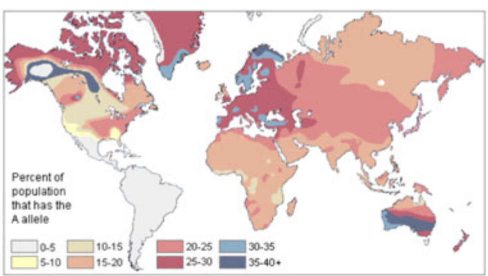

Figure 9. Distribution of type A blood (www.anthro.palomar.edu/vary/vary_3.htm).

three types of family relations, among which the blood typing process can be conducted simultaneously among independent families. Meanwhile, a decentralised algorithm is designed for that in intermarriage and hierarchical families. We conduct extensive simulations to test the performance of the proposed methods, which also validate the effects of the impact factors, including the number of children, the order of blood typing, the probability of correct estimation and the blood type distribution. For the future work, we are interested in applying the proposed method into practice, and predicting the blood type distribution of the population in a certain area, given that of the current generation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported by NSF [grant number ECCS 1231461], [grant number ECCS 1128209], [grant number CNS 1138963], [grant number CNS 1065444] and [grant number CCF 1028167]; and NSFC [grant number 61272151], [grant number 61472451]; ISTCP [grant number 2013DFB10070]; the China Hunan Provincial Science and Technology Program [grant number 2012GK4106]; and the Chinese Fundamental Research Funds for the Central Universities [grant number 531107040845].

Notes

1. Email: jiewu@temple.edu
2. Email: csgjwang@csu.edu.cn
3. Email: huanyang.zheng@temple.edu

References

[1] S. Amer-Yahia, A. Doan, J. Kleinberg, N. Koudas, and M. Franklin, *Crowds, clouds, and algorithms: Exploring the human side of big data applications*, Proc. ACM SIGMOD (2010), pp. 1259–1260.
[2] P.A. Bernstein, V. Hadzilacos, and N. Goodman, *Concurrency Control and Recovery in Database Systems*, Addison Wesley, Reading, MA, USA, 1987.
[3] M.S. Bernstein, G. Little, R.C. Miller, B. Hartmann, M.S. Ackerman, D.R. Karger, D. Crowell, and K. Panovich, *Soylent: A word processor with a crowd inside*, Proc. ACM UIST (2010), pp. 313–322.
[4] C. Callison-Burch, *Fast, cheap, and creative: Evaluating translation quality using Amazon's mechanical turk*, Proc. EMNLP (2009), pp. 286–295.
[5] Y. Chon, N.D. Lane, F. Li, H. Cha, and F. Zhao, *Automatically characterizing places with opportunistic crowdsensing using smartphones*, Proc. ACM UbiComp (2012), pp. 481–490.
[6] K.M. Cramer, and E. Imaike, *Personality, blood type, and the five-factor model*, Pers. Individual Diff. 32(4) (2002), pp. 621–626.
[7] M.J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin, *CrowdDB: Answering queries with crowdsourcing*, Proc. ACM SIGMOD (2011), pp. 61–72.
[8] M.J. Franklin, B. Trushkowsky, P. Sarkar, and T. Kraska, *Crowdsourced enumeration queries*, Proc. IEEE ICDE (2013), pp. 673–684.
[9] R.K. Ganti, F. Ye, and H. Lei, *Mobile crowdsensing: Current state and future challenges*, IEEE Commun. Mag. 49(11) (2011), pp. 32–39.
[10] J. Gao, X. Liu, B. Ooi, H. Wang, and G. Chen, *An online cost sensitive decision-making method in crowdsourcing systems*, Proc. ACM SIGMOD (2013), pp. 217–228.
[11] C.M. Grinstead, and J.L. Snell, *Introduction to Probability*, 2nd ed., American Mathematical Society, Providence, RI, USA, 1997.
[12] S. Guo, A. Parameswaran, and H. Garcia-Molina, *So who won?: Dynamic max discovery with the crowd*, Proc. ACM SIGMOD (2012), pp. 385–396.

[13] M. He, B. Wolpin, K. Rexrode, J.E. Manson, E. Rimm, F.B. Hu, and L. Qi, *ABO blood group and risk of coronary heart disease in two prospective cohort studies*, Arterioscler. Thromb. Vasc. Biol. (2012).
[14] C. Hu, B.B. Bederson, and P. Resnik, *Translation by iterative collaboration between monolingual users*, Proc. HCOMP ACM KDD Workshop Hum. Comput. (2010), pp. 54–55.
[15] W. Jiang, G. Wang, and J. Wu, *Generating trusted graphs for trust evaluation in online social networks*, Fut. Gen. Comp. Syst. 31 (2014), pp. 48–58.
[16] D. Lester, and J. Gatto, *Personality and blood group*, Pers. Individual Diff. 8(2) (1987), p. 267.
[17] A. Papoulis, *Bayes' Theorem in Statistics and Bayes' Theorem in Statistics (Reexamined). §3-5 and 4-4*, in *Probability, Random Variables, and Stochastic Processes*, 2nd ed. Mcgraw-Hill, New York, 1984, pp. 38–39, 78–81, and 112–114.
[18] M. Rogers, and A. Ian Glendon, *Blood type and personality*, Pers. Individual Diff. 34(7) (2003), pp. 1099–1112.
[19] *Teaching computers to read: Google Acquires Recaptcha*, Google Retrieved, 2009.
[20] J. Wang, G. Li, T. Kraska, M.J. Franklin, and J. Feng, *Leveraging transitive relations for crowdsourced joins*, Proc. ACM SIGMOD (2013), pp. 229–240.
[21] D. Yang, G. Xue, X. Fang, and J. Tang, *Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing*, Proc. Mobicom (2012), pp. 173–184.
[22] T. Yan, V. Kumar, and D. Ganesan, *Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones*, Proc. ACM MobiSys (2010), pp. 77–90.