Aggregating Expert Opinions in Support of Medical Diagnostic Decision-Making

Clemens Gangl[1], Jan Maly[2], Martin Lackner[2], and Stefan Woltran[2]

[1] Department of Internal Medicine II,
Medical University of Vienna, Austria
clemens.gangl@meduniwien.ac.at

[2] Databases and Artificial Intelligence Group,
TU Wien, Vienna, Austria
{maly,lackner,woltran}@dbai.tuwien.ac.at

Abstract. Medical doctors are often faced with challenging diagnostic decisions, which require the consideration of all eligible differential diagnoses. Diagnostic decisions (i.e., which test to order next in a given situation) have a high impact, as non-targeted diagnostic strategies may cause delayed treatments. It is thus desirable for medical professionals to be able to tap into the knowledge from more experienced colleagues - a process which can be fostered and supported by knowledge-based software tools. In this position paper, we outline the potential and challenges of applying methods from Computational Social Choice (COMSOC) to aggregate expert advice on diagnostic strategies.
Gathering and aggregating expert opinions is a challenging task, especially in the medical domain. We discuss the necessary requirements for COMSOC methods to be applicable in the diagnostic support setting, in particular requirements for opinion elicitation and opinion aggregation. The main goal of our research is to build a system that supports diagnostic decision-making based on reliable expert knowledge. Principled methods and analyses from COMSOC guarantee that recommendations are reliable, sound, and explainable.

Keywords: Medical Decision Support · Computational Social Choice · Preference Aggregation · Preference Elicitation · Multi-Winner Voting

1 Vision and Challenges

Medical diagnostic decision-making is becoming an increasingly demanding task: diagnostic options are gaining in complexity while higher efficiency is required. In a nutshell, more patients need to be diagnosed more accurately in less time. In their daily routine, medical doctors are thus often faced with challenging clinical situations that require immediate but solid diagnostic decisions.

A medical diagnosis usually results out of an iterative process, starting with a structured interview (anamnesis) to gather descriptive information on characteristics and course of the complaints, the medical history, etc., followed by a physical exam, often supplemented by laboratory tests, imaging methods or even invasive diagnostics. Extensive scientific effort is directed on developing strategies how to deduce reasonable, probability-ranked differential diagnosis from these information to provide physicians a profound basis for their further diagnostic decision-making. Conventional rule-based systems are nowadays complemented by machine learning approaches, which extract knowledge from electronic health reports, case-reports and other literature [1, 2, 11].

Given a set of potential differential diagnoses, physicians then need to decide what would be the optimal next test to further narrow down the list or to finally confirm one of the possible diagnoses. This decision has to be based on various considerations such as the hazardousness, availability, or time requirements of diagnostic options. Most importantly, non-targeted diagnostic strategies may cause unnecessary tests and imperfect results, leading to delayed treatments or misdiagnoses [12]. In practice, a doctor will determine which diagnostic decisions are most sensible based on many trade-offs between hard-to-quantify variables and personal experience. Given the high stakes of diagnostic decisions, medical professionals occasionally seek to back up their reasoning by consulting more specialized or experienced colleagues. However, this reassurance is not always possible due to constraints in the available amount of both time and effort.

In this position paper, we report on challenges in the aggregation of expert opinions that arise in the development of a system to support diagnostic decisions. The focus is on collecting and aggregating opinions on the most reasonable next test an expert would perform in a given clinical situation but not on rating potential differential diagnoses. Such a system should provide advice by listing the most reasonable further diagnostic options based on available observations. In particular, we envision an application in emergency medicine, with a preliminary focus on the frequent chief complaints *shortness of breath* and *chest pain* combined with their related differential diagnoses and other connected findings.

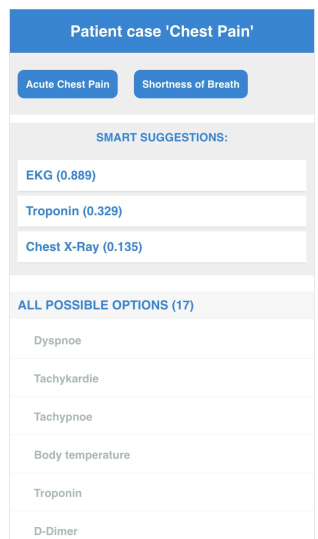

Fig. 1: Exemplary application screen

In a typical use case, a doctor examines a patient and is unsure whether they considered all potential differential diagnoses. As a consequence, they are unsure about choosing the next diagnostic step and would like to reassure their diagnostic decision by consulting the proposed system. Using a mobile device (cf. Figure 1), the doctor states the patient's chief complaint (e.g. shortness of breath). The system will then ask for further related signs and symptoms (e.g. fever, chest pain, exhaustion). Based on this information, it will query a database containing collected advice for this and related situations. The system lists recommendable next diagnostic steps (e.g. lab tests or imaging procedures) by aggregating available expert opinions.

The aggregation of expert opinions is the conceptual backbone of our system. It is paramount to give only solid, trustworthy advice bolstered by expert opinions. Our system builds on principled methods from research in Computational Social Choice (COMSOC) [4], which allows us to explain recommendations and to give guarantees on their soundness.

This system could achieve an increase in medical decision quality, reducing the time and necessary effort required to find correct diagnoses and therefore reduce costs and potentially harmful diagnostic procedures while medical professionals maintain full autonomy in their decisions at the same time.

Challenges. In this paper, we address in particular the following challenges:

1. *A vast number of possible diagnostic options:* To reduce this number, we generate a shortlist by using a knowledge base of common medical knowledge, which is sufficient to rule out clearly irrelevant diagnostic options given a patient's current condition.
2. *Trustworthiness and reliability*: To ensure the acceptance of medical professionals, any given advice is based on expert opinions, which is elicited and aggregated by mathematically principled and sound methods from COMSOC.
3. *Collecting expert opinions*: Acquiring advice from experts is a non-trivial task, especially in the medical domain. Specific difficulties are the high cost of data acquisition, the large number of possible diagnostic options, and the recognition of similar situations.
4. *Inconsistency of expert opinions*: Since expert opinions may be conflicting, a solid aggregation mechanism has to deal with unclear situations. Sometimes it is necessary to refrain from giving a recommendation; if a recommendation is given, it must always be possible to explain this recommendation.
5. *Availability and accessibility.* Diagnostic decisions often need to be made while examining the patient ("point of care decisions"). Hence, the end users' interface has to be usable with low effort, suggestions should be easily comprehensible, and the system has to be accessible everywhere and immediately.

Vision. An accessible system that provides diagnostic advice can achieve an increase in medical decision quality, reduce the time and necessary effort required to find correct diagnoses, and therefore reduce costs and potentially harmful diagnostic procedures. The proposed system does not diminish the full autonomy of medical professionals, but provides fast and reliable access to advice from experienced specialists.

2 Preselection of Feasible Diagnostic Options

In general, a patient's current condition can be summarized as a *situation* which represents a combination of *findings* like signs and symptoms, laboratory values and results of machine aided examinations (gained from EKGs or CT scans for example). In addition, *preconditions* (e.g. the patient's age or if the patient is an organ recipient) can be linked to a defined *situation* as this is additional, valuable information.

Figure 1 shows an exemplary application screen after the user (doctor) stated 'acute chest pain' and 'shortness of breath' as chief complaints in a first step. Based on these stated findings, the system can internally compile an elementary list of possible differential diagnoses by a combinatorial approach using a knowledge base which maps diseases to all possible related findings.

Unfortunately, this often results in a large number of possible diagnostic options, too many to present to end users in a meaningful way. It is thus necessary to further filter these generated diagnostic options based on expert opinions.

3 Aggregating Expert Opinions

Computational Social Choice offers many methods for the aggregation of (possibly contradictory) opinions, which are generally referred to as voting rules. Voting rules usually take as input a collection of preferences and output either a single winner or a set of winners. Multiwinner approval voting is often considered the best choice for shortlisting, i.e., preselecting a group of most sensible options [3, 6, 9]. Most research on shortlisting is focused on multiwinner voting with a fixed number of winners [7]. A clinical decision support system as we envision should not pick a fixed number of options, but rather list *all* options that deserve a doctor's consideration. Therefore, we focus on multiwinner voting rules with a variable number of winners, which can be formalized as follows:

An approval-based election $E = (C, V)$ consists of a set of candidates $C = \{c_1, \dots, c_m\}$ and a voter profile $V = \{v_1, \dots, v_n\}$ where $v_i \subseteq C$ is the set of candidates approved by voter i. The *approval score* $a_E(c_i)$ of a candidate i in election E is the number of approvals of candidate c_i in V. Let E be an election and $S \subseteq C$ be a set of candidates. Then we call $a_E(S) := \Sigma_{c \in S} a_E(c)$ the approval of S. Furthermore, we call $na_E(S) := \Sigma_{c \in S} 2a_E(c) - m$ the net-approval of S. Finally, we write $S \in \mathcal{W}_E^r$ if S is a possible set of winners under a voting rule r.

For shortlisting, it is natural to assume that no winner can be (strictly) less approved than a non-winner, i.e. if $a_E(c_i) > a_E(c_j)$ and $c_j \in W_E^r$ for some $W_E^r \in \mathcal{W}_E^r$ then also $c_i \in W_E^r$. If a voting rule satisfies this property, called *efficiency*, the problem of determining the winners reduces to where to draw the line between winners and non-winners[3]. Figure 2 gives an overview over voting rules considered in the literature that satisfy efficiency.

Majority voting, i.e. fixing some threshold of approval for being a winner—for example 50%—is the simplest voting rule that satisfies efficiency. Other voting rules, like the *Next-k rule* and *First majority* take into account the approval of either all other candidates or at least of the neighboring candidates. In [9], Kilgour proposed the voting rules CSA and NCSA that take the size of the set of winners into account. In [8], Faliszewski et al. showed that CSA and NCSA mostly choose committees of size one. They proposed generalized versions q-CSA

[3] Additionally, some tie-breaking may occur.

Majority voting A candidate is a winner if and only if he has more than 50% approval, i.e. $\mathcal{W}_E^r = \{W_E^r\}$ and $c_i \in W_E^r$ iff $a_E(r) > \frac{n}{2}$.

Next-k rule [3] The cut between winners and non-winners is made if a candidate has more approval than the next k candidates together. This can be formalized as follows: Let $c_1, \ldots, c_m$ be an enumeration[a] $a_E(c_i) \leq a_E(c_{i-1})$. Then, $c_i \in W_E^r$ iff $a_E(c_{i-1}) \leq \Sigma_{j=0}^{k-1} a_E(c_{i+j})$.

First majority [9] The smallest sets of candidates that have more than 50% of all approvals are the possible sets of winners. This can be formalized as follows: Let $c_1, \ldots, c_m$ be an enumeration[b] of the candidates such that $a_E(c_i) \leq a_E(c_{i-1})$. Then $c_i \in W_E^r$ iff $\Sigma_{j<i} a_E(c_j) < \Sigma_{i \leq k \leq m} a_E(c_k)$.

Capped Satisfaction Approval Voting (q-CSA) [8] The sets of candidates with the best approval to size ratio are the possible sets of winners, i.e. $S \in \mathcal{W}_E^r$ for $S \subset C$ if $\frac{a_E(S^*)}{|S^*|^q} \leq \frac{a_E(S)}{|S|^q}$ holds for all $S^* \in \mathcal{P}(C)$.

Net Capped Satisfaction Approval Voting (q-NCSA) [8] The sets of candidates with the best net approval to size ratio are the possible sets of winners, i.e. $S \in \mathcal{W}_E^r$ for $S \subset C$ if $\frac{na_E(S^*)}{|S^*|^q} \leq \frac{na_E(S)}{|S|^q}$ holds for all $S^* \in \mathcal{P}(C)$.

[a] Every enumeration with this property gives the same set of winners.

[b] If different enumeration with this property exist, we get a different set of winners for every enumeration. All of them are possible sets of winners.

Fig. 2: Efficient voting rules

and q-NCSA where 1-CSA and 1-NCSA are equivalent to the original versions and showed empirically that these perform significantly better if an adequate value for the parameter $q \in [0, 1]$ is chosen.

There are several further desiderata for a voting rule that could be considered in the diagnostic support setting: (1) The voting rule should not discount any options that are approved by a significant majority of experts. (2) The voting rule should not select any winners that are not approved by significant percentage of the experts. (3) If the approval of two options does not differ significantly, either both or neither option should be a winner. (4) The set of winners should represent a significant fraction of available opinions. Our first research goal will be to formalize these desiderata.

Main Research Objective 1 *Define axioms that capture the desiderata for a voting rule in the diagnostic support setting.*

Once these axioms are fixed, we want to determine if there exists a voting rule that satisfies all desirable axioms and if not to find an acceptable comprise.

Main Research Objective 2 *Find a voting rule that satisfies as many desirable axioms as possible.*

Depending on the exact formulation of the axioms, it seems probable that it is possible to satisfy desiderata (1) and (2) with a rule similar to *Majority* *voting*, desiderata (3) with a rule similar to the *Next-k rule* and desiderata (4) with a version of *First majority*. Further research is needed to see if it is possible to do better.

4 Efficient Opinion Elicitation

Acquisition of expert knowledge is expensive. Consequently, it is of high importance to acquire knowledge as efficiently as possible. In particular, the goal is to pose questions that yield the highest information gain, and to ask as few questions as possible without compromising recommendation quality.

Main Research Objective 3 *Find opinion elicitation methods that* provide the necessary information *to successfully compute a chosen opinion aggregation method with* minimal information requirements*.*

In general, it is not necessary to know the opinion of all experts on all diagnostic options. Some options will have sufficiently strong support to clearly include them in the recommendation; others can soon be rejected. This requires that we adapt the opinion aggregation methods of Section 3 to be able to deal with incomplete information. Incomplete information increases the uncertainty in the aggregation process and hence increases the likelihood of situations where no recommendation can be made with sufficient confidence. Opinion elicitation methods thus have to actively identify and avoid such unclear situations by asking questions that settle uncertainties. One of the most promising elicitation strategies discussed in the literature is vote elicitation based on minmax-regret [10]. While this method was developed for preferential, single-winner voting, it can be adapted to our approval voting setting. Another common approach is to run the elicitation until the remaining uncertainties are inconsequential (the winning candidates are fully determined [5]). This approach is not desirable in our case, as giving *some* trustworthy recommendations is preferable to waiting until *all* available options are sufficiently evaluated.

Certain limitations also follow from practical considerations. For example, it is unreasonable to expect experts to select recommendations from huge lists of options; some preselection is necessary. The elicitation procedure has to follow such practical guidelines. A positive aspect are interrelations between diagnostic options: it may be possible to group related options (e.g., imaging methods) to quickly narrow down the full list.

5 Discussion and Outlook

In this position paper, we presented fundamental research questions in Computational Social Choice, motivated by an application in diagnostic decision-making. The goal is to tap into expert knowledge and give recommendations based on this knowledge. In contrast, we do not wish to rate potential differential diagnoses but rather our approach can complement established clinical decision support systems by recommending the next best diagnostic test in given situations.

To this end, we proposed to use principles of multi-winner voting and discussed how to adapt them to fit the needs of this particular application domain. An axiomatic analysis, as sketched in this paper, provides a solid basis for the selection of suitable multi-winner voting rules. It is evident that in high-stake domains—in our case medical decision-making—it is essential to use reliable and explainable mechanisms. We believe that methods from COMSOC can rise to this challenge, but further research is required to design suitable mechanisms specifically for our intended application scenario.

Acknowledgments This work was supported by the Austrian Science Foundation FWF, grants P31890 and Y698.

References

1. Berner, E.S.: Clinical decision support systems, vol. 233. Springer (2007)
2. Bond, W.F., Schwartz, L.M., Weaver, K.R., Levick, D., Giuliano, M., Graber, M.L.: Differential diagnosis generators: An evaluation of currently available computer programs. Journal of General Internal Medicine **27**(2), 213–219 (2012). https://doi.org/10.1007/s11606-011-1804-8
3. Brams, S.J., Kilgour, D.M.: Narrowing the field in elections: the next-two rule. Journal of Theoretical Politics **24**(4), 507–525 (2012)
4. Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D.: Handbook of Computational Social Choice. Cambridge University Press (2016)
5. Conitzer, V., Sandholm, T.: Vote elicitation: Complexity and strategy-proofness. In: Proc. of AAAI'02. pp. 392–397. AAAI Press (2002)
6. Elkind, E., Faliszewski, P., Skowron, P., Slinko, A.: Properties of multiwinner voting rules. Social Choice and Welfare **48**(3), 599–632 (Mar 2017)
7. Faliszewski, P., Skowron, P., Slinko, A., Talmon, N.: Multiwinner voting: A new challenge for social choice theory. In: Endriss, U. (ed.) Trends in Computational Social Choice, chap. 2, pp. 27–47. AI Access (2017)
8. Faliszewski, P., Slinko, A., Talmon, N.: The complexity of multiwinner voting rules with variable number of winners. CoRR **abs/1711.06641** (2017)
9. Kilgour, D.M.: Approval elections with a variable number of winners. Theory and Decision **81** (02 2016)
10. Lu, T., Boutilier, C.: Robust approximation and incremental elicitation in voting protocols. In: Proc. of IJCAI'11. pp. 287–293. IJCAI'11, AAAI Press (2011)
11. Riches, N., Panagioti, M., Alam, R., Cheraghi-Sohi, S., Campbell, S., Esmail, A., Bower, P.: The effectiveness of electronic differential diagnoses (DDX) generators: A systematic review and meta-analysis. PLoS ONE **11**(3), 1–26 (2016)
12. Schiff, G.D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B.L., Elstein, A.S., Hasler, S., Kabongo, M.L., Krosnjar, N., et al.: Diagnostic error in medicine: Analysis of 583 physician-reported errors. In: Archives of Internal Medicine. pp. 1881–7. JAMA Internal Medicine (2009)